MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Dec. 2005, p. 608–634 Vol. 69, No. 4
1092-2172/05/$08.00+0 doi:10.1128/MMBR.69.4.608–634.2005
Copyright © 2005, American Society for Microbiology. All Rights Reserved.

# Comparative Genomic Analyses of the Bacterial Phosphotransferase System

Ravi D. Barabote and Milton H. Saier, Jr.*

*Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093-0116*



## INTRODUCTION

Four decades ago, Kundig et al. reported the discovery of a novel sugar-phosphorylating system in *Escherichia coli* (38). The unique features of this phosphotransferase system (PTS) included the use of phosphoenolpyruvate (PEP) as the phosphoryl donor for sugar phosphorylation and the presence of three essential catalytic entities, termed enzyme I, enzyme II, and HPr (heat-stable, histidine-phosphorylatable protein). The discovery of this system provided an explanation for pleiotropic carbohydrate-negative mutants of *E. coli* described as early as 1949 (20).

In 1964, the three recognized activities of the PTS were presumed to correspond merely to three proteins. We now know that dozens of PTS proteins are present in the *E. coli* cell and that thousands of PTS protein homologues occur in other bacteria. Numerous genes encoding these proteins have been fully sequenced, and their phylogenetic relationships have been described (29, 81; Nguyen et al., unpublished data).

The bacterial PTS catalyzes the concomitant transport and phosphorylation of its sugar substrates (62, 96). It is a complex system that consists of general cytoplasmic energy-coupling proteins, enzyme I (EI) and HPr, which lack sugar specificity, and membranous enzyme II complexes, each specific for one or a few sugars. The enzyme II complexes usually consist of three proteins or protein domains, namely, IIA, IIB, and IIC. However, the enzyme II complexes of one of the PTS families, the mannose family, have one additional membrane-spanning protein or domain, called IID (72). Phosphoryl relay proceeds sequentially from PEP to EI, HPr, IIA, IIB, and finally, the incoming sugar, which is transported across the membrane via the integral membrane IIC porter (Table 1 and Fig. 1A).

When the PTS was discovered, a single function was recognized, namely, sugar phosphorylation. Forty-one years later, we find that this system plays roles in many surprising aspects of bacterial physiology. Established primary functions of the system include sugar reception, transport, and phosphorylation, whereas secondary functions include a variety of ramifications for metabolic and transcriptional regulation (Table 2) (37, 39, 84, 86, 87, 99, 100; for reviews, see references 62, 76, 79, and 84).

Genetic evidence has indicated that in various bacteria, processes regulated by the PTS include (i) transport of non-PTS carbon sources (35), (ii) the net production of carbon and energy storage sources, such as poly-β-hydroxybutyrate (64, 90) and glycogen (91), (iii) the switch between fermentative and respiratory metabolism (36), (iv) flagellar motility (54), and (v) the control of $\sigma^{54}$-dependent transcription of carbon and nitrogen metabolic genes (12, 46, 63, 70, 74). Some evidence implicates the PTS in

* Corresponding author. Mailing address: Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093-0116. Phone: (858) 534-4084. Fax: (858) 534-7108. E-mail: msaier@ucsd.edu.

TABLE 1. Structural complexity of PTS

| Protein or process | Description |
|---|---|
| IIC | Permease and receptor (sugar specific) |
| IIB | Direct phosphoryl donor (permease specific) |
| IIA | Indirect phosphoryl donor (family specific) |
| IID | Mannose family-specific auxiliary protein (essential but of unknown biochemical function) |
| EI and HPr | General energy-coupling proteins (PTS pathway specific) |
| Enolase | Upstream energy-yielding enzyme (PEP generating) |
| Phosphoglucoisomerase | Downstream substrate-converting enzyme |
| Glycolysis | Interconnecting cyclic pathway |
| PTS + glycolysis | Metabolite-induced metabolon |

TABLE 2. Functional complexity of PTS

| PTS function |
|---|
| Chemoreception |
| Transport |
| Sugar phosphorylation |
| Protein phosphorylation |
| Regulation of non-PTS sugar transport and metabolism |
| Regulation of carbon metabolism |
| Regulation of carbon storage |
| Regulation of fermentation versus respiration |
| Regulation of cellular motility |
| Coordination of nitrogen and carbon metabolism |
| Regulation of non-carbon-compound transport |
| Regulation of gene expression |
| Regulation of pathogenesis |
| Regulation of cell physiology |
| Regulation of cell division |

the regulation of cell division (44, 63). Still other regulatory functions of the PTS and their physiological consequences are listed in Table 2.

In *Escherichia coli*, paralogues of EI, HPr, and the fructose IIA protein ($IIA^{Fru}$), designated nitrogen enzyme I ($EI^{Ntr}$), nitrogen HPr (NPr), and nitrogen IIA protein ($IIA^{Ntr}$), respectively, constitute a phosphoryl transfer chain (Fig. 1A) that has been shown to exhibit little enzymatic cross-reactivity with the classical sugar-transporting phosphoryl transfer chain consisting of EI, HPr, and various sugar-specific enzyme II complexes (65). This nitrogen-related phosphoryl transfer chain presumably functions only in regulation (46, 47, 63). $EI^{Ntr}$ homologues have been shown to cluster phylogenetically together, distantly from all other enzyme I homologues (29).

Phylogenetic data have shown that NPr in *E. coli* is a distant paralogue of HPr (29, 63). Sequence characteristics that distinguish NPr from HPr as well as $EI^{Ntr}$ from EI have been described previously (65). $EI^{Ntr}$ consists of two domains, an N-terminal domain of 127 amino acids homologous to the N-terminal "sensory" domain of the NifA protein of *Azotobacter vinelandii* (2) and a C-terminal domain of 578 amino acids homologous to all currently sequenced enzyme I proteins. $EI^{Ntr}$ may serve a sensory function linking carbon and nitrogen metabolism (69). A mutation of the orthologous $EI^{Ntr}$-encoding *ptsP* gene of *A. vinelandii* resulted in impaired metabolism of poly-β-hydroxybutyrate as well as diminished respiratory protection of nitrogenase under carbon-limiting conditions (90). In addition to $EI^{Ntr}$ and NPr, *E. coli* encodes within its genome three additional EI paralogues and four additional HPr paralogues. The functions of most of these proteins are still unknown.

The biochemical detection of novel PTS proteins in bacteria as diverse as *Ancalomicrobium adetum* (85), *Spirochaeta aurantia*

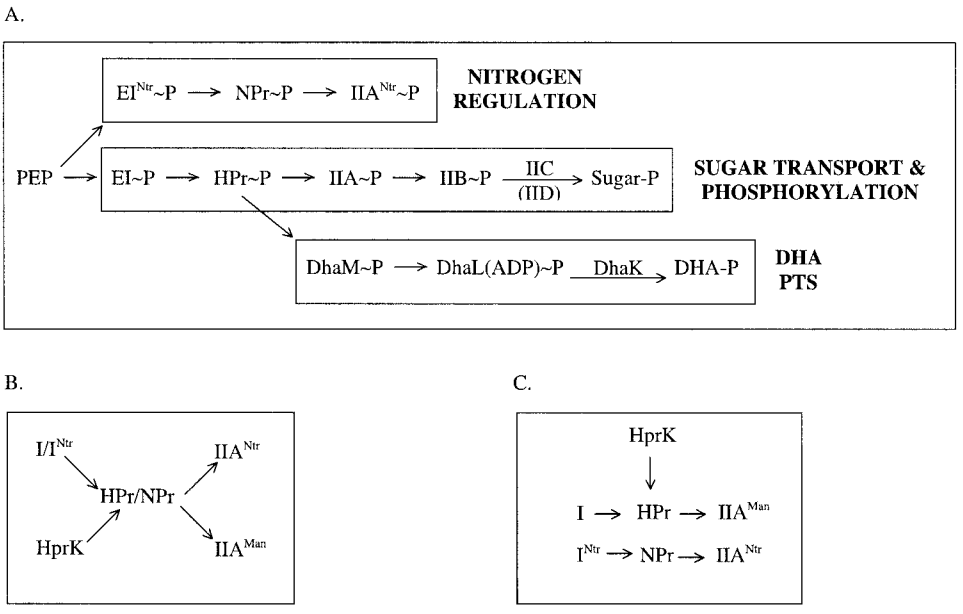



FIG. 1. Phosphoryl transfer chains of the bacterial phosphotransferase system. (A) Various phosphoryl relay chains of the PTS. (B) Proposed regulatory phosphoryl transfer chains based on the data in Table 7. (C) Putative phosphoryl chains in *Bradyrhizobium japonicum*.

TABLE 3. Sugar substrates transported by various PTS permeases[a]

| Family (TC no.)[b] | Subfamily (TC no.)[b] | Substrates |
|---|---|---|
| Glc (4.A.1) | Glucose (4.A.1.1) | Glucose, *N*-acetylglucosamine, maltose, glucosamine, and α-glucosides |
| | Glucoside (4.A.1.2) | β-Glucosides, sucrose, trehalose, and *N*-acetylmuramic acid |
| Fru (4.A.2) | | Fructose, mannitol, mannose, and 2-*O*-α-mannosyl D-glycerate |
| Lac (4.A.3) | Lactose (4.A.3.1) | Lactose, aromatic β-glucosides, and cellobiose |
| | *N,N′*-diacetylchitobiose (4.A.3.2) | *N,N′*-Diacetylchitobiose and lichenan oligosaccharides |
| Gut (4.A.4) | | Glucitol (sorbitol) and 2-methyl-D-erythritol |
| Gat (4.A.5) | | Galactitol and D-arabitol |
| Man (4.A.6) | | Glucose, mannose, sorbose, fructose, glucosamine, galactosamine, *N*-acetylgalactosamine, and several other sugars; broad-specificity porters |
| Asc (4.A.7) | | L-Ascorbate (anaerobic utilization) |

[a] Sugars transported by functionally characterized members of the seven PTS permease families are indicated.
[b] The transporter classification number (TC no.) for each of the PTS families and subfamilies is provided. For more information, refer to http://www.tcdb.org/.

(83), *Acholeplasma laidlawii* (28), *Listeria monocytogenes* (50), and several antibiotic-producing species of *Streptomyces* (6, 104) suggests the involvement of PTS proteins in cellular processes distinct from those currently recognized. It is worth noting that other families of transport systems, such as the family of ATP-binding cassette (ABC)-type permeases (27) and the major facilitator superfamily (55), apparently do not participate in metabolic and transcriptional regulation to the extent observed for the PTS.

Recently, a nontransporting enzyme II complex was characterized in *E. coli* that phosphorylates dihydroxyacetone (DHA) at the expense of PEP, using three soluble DHA-specific proteins in addition to EI and HPr (25). The three components of the DHA enzyme II complex are designated DhaK, DhaL, and DhaM. The DhaM protein of *E. coli* is a tridomain protein consisting of an N-terminal $IIA^{Dha}$ domain that is distantly related to $IIA^{Man}$, a central HPr domain, and a C-terminally truncated EI domain. All three domains have been shown to be phosphorylated, using PEP and the classical enzyme I and HPr proteins as phosphoryl donors. The various currently recognized phosphoryl relay chains of the PTS are depicted in Fig. 1A.

A bifunctional HPr kinase/phosphorylase (HprK) that catalyzes the phosphorylation of HPr at Ser-46 at the expense of ATP (18, 53, 66) as well as the dephosphorylation of P-Ser-HPr (49) was discovered in *Streptococcus pyogenes* (19). It plays an important regulatory role in at least three different cellular processes: (i) sugar uptake via the PTS, (ii) catabolite control protein A (CcpA)-mediated carbon catabolite repression, and (iii) inducer control via expulsion and exclusion mechanisms (for reviews, see references 7, 78, and 79). HprK homologues from gram-positive as well as gram-negative bacteria have been proposed to carry out different functions (7, 29, 97). Detailed phylogenetic analyses of HprK homologues from several bacteria have recently been presented (97).

The currently recognized structural and functional complexity of the PTS is impressive (Tables 1 and 2), but the available evidence suggests that its functional ramifications are only now beginning to be realized (7, 23). Only half of the PTS proteins recognized in *E. coli* have been functionally characterized (101), and as revealed by the present genomic analyses, almost all PTS proteins in other organisms are uncharacterized.

Based on the phylogeny of the IIC proteins, seven PTS permease families are currently recognized, namely, the (i) glucose (including glucoside) (Glc), (ii) fructose (including mannitol) (Fru), (iii) lactose (including *N,N*-diacetylchitobiose) (Lac), (iv) galactitol (Gat), (v) glucitol (Gut), (vi) mannose (Man), and (vii) L-ascorbate (Asc) families. Various sugar substrates transported by the few functionally characterized members of each of these families are presented in Table 3.

Proposed evolutionary origins of the different PTS families have been discussed (81). Briefly, the substrate-recognizing protein constituents of the PTS (enzymes IIC) are derived from at least four independent sources. Some of the non-PTS precursor constituents have been identified, and the evolutionary pathways taken have been proposed (81). Analyses suggest that two of these independently evolving systems (Gat and Dha) are still in transition, as they have not yet acquired the full-fledged characteristics of PTS enzyme II complexes. The mosaic nature of PTS enzyme II complexes has also been documented (81).

## OVERVIEW OF GENOME ANALYSES

The last few years have witnessed an explosion in the amount of sequence data available for analysis. Genome sequencers sometimes deposit new sequences into databases without rigorous scrutiny. Incorrect or imprecise annotations of genes and gene products not only obscure important genomic information but also lead to the proliferation of erroneous annotations of other genomes. Frequently, sequencing errors conceal important open reading frames (ORFs), and distant phylogenetic relationships are not noticed. Here we report our computational analyses of several completely sequenced genomes.

A total of 202 genomes were screened for the presence of homologues of all currently known constituents of the bacterial phosphotransferase system. These proteins include the general energy-coupling proteins EI and HPr, the IIA, IIB, IIC, and IID constituents of the enzyme II complexes, the DHA PTS proteins (DhaM, DhaL, and DhaK), and HprK. We used criteria established by Bächler (3) to detect homologues of the *E. coli* dihydroxyacetone PTS proteins (25). GenBank sequence identification (GI) numbers are provided for the proteins discussed in the text.

All completely sequenced genomes were obtained from NCBI (ftp://ftp.ncbi.nih.gov/genomes/). A standalone version (release 2.0.10) of BLAST (1) was obtained ftp://ftp.ncbi.nih.gov/blast/ and was employed to identify PTS homologues, us-

TABLE 4. Overview of PTS analyses of whole genomes

| Parameter | No. |
|---|---|
| Total genomes analyzed[a] | 202 |
| Bacterial genomes[b] | 174 |
| Archaeal genomes | 19 |
| Eukaryotic genomes | 9 |
| Total different bacterial species analyzed[c] | 136 |
| Bacterial species with no PTS-encoding genes | 30 |
| Bacterial species that encode homologues of PTS proteins | 106 |
| Bacterial species lacking PTS permeases | 29 |
| Bacterial species encoding at least one complete PTS permease | 77 |
| Total archaeal species analyzed[c] | 19 |
| Archaea with no PTS homologues | 19 |
| Total eukaryotic species analyzed | 9 |
| Eukaryotes with no PTS homologues | 9 |

[a] All of the archaeal and eukaryotic genomes analyzed belong to separate genera. However, the bacterial genomes analyzed encompass 136 different species that belong to 89 genera.

[b] Of the 174 bacterial genomes analyzed, 60 of the strains belong to 22 different bacterial species. Of these, 51 strains that belong to 18 different bacterial species encode PTS protein homologues.

[c] Only completely sequenced genomes were analyzed. Incompletely sequenced genomes (e.g., sequenced without closure) were not included in the present analyses.

ing 57 functionally characterized PTS proteins obtained from the transporter classification database (TCDB; http://www.tcdb.org) as query sequences against each of the completely sequenced genomes. Over 3,000 protein homologues were retrieved and analyzed. Previous annotations were ignored, and all proteins were reexamined. Protein domains were identified and assigned using the NCBI CD-Search tool (42). Such rigorous scrutiny often revealed errors in previous annotations.

Enzyme II complexes that include IIA, IIB, and IIC (as well as $IID^{Man}$ in the case of the mannose family) were considered to constitute complete PTS permeases. Enzyme IIC proteins that lack a cognate IIA and/or IIB domain/protein were considered incomplete systems. It should be noted that in the Glc family, and only the Glc family, several PTS porters have the IIB and IIC domains fused but lack their own IIA domain. Instead, these systems use the $IIA^{Glc}$ protein of another system (101). Such Glc-type systems were considered complete.

Table 4 presents an overview of our PTS genome analyses. Except for phosphoenolpyruvate synthases and pyruvate:phosphate dikinases, which are distantly related to EIs of the PTS (101) and are present in both eukaryotes and archaea, and except for a single HPr kinase homologue found in an archaeon, *Methanopyrus kandleri* (97), no homologues of PTS proteins were identified in any eukaryote or archaeon. Within the bacterial domain, 22% of the species analyzed (30 organisms) encode no recognizable PTS protein homologues. Twenty-one percent (29 organisms) encode cytoplasmic PTS phosphoryl transfer proteins but lack complete membrane-integrated PTS enzyme II complexes. Finally, 57% of these bacterial species (77 organisms) have at least one complete PTS enzyme II complex as well as the requisite PTS energy-coupling proteins.

## RELATIVE DISTRIBUTION OF PTS PERMEASE TYPES

The complete PTS permeases identified in distinct bacterial species were tabulated according to the family to which they belong, as classified in the TCDB (10, 77). The greatest representation was observed for the Glc family (30%), with the Fru family (25%) coming in second place (Fig. 2). The order of occurrence of the members of the seven families was Glc (30%) > Fru (25%) > Man (15%) > Lac (14%) > Asc (9%) > Gat (4%) > Gut (3%). The Glc, Fru, Man, and Lac systems often occur in multiple copies, while the Asc, Gat, and Gut systems are usually found as a single copy per organism (Fig. 2). Furthermore, about equal proportions of organisms possess the Fru and Glc systems (86% and 83%, respectively), followed by the Asc (47%), Lac (44%), Man (42%), Gut (18%), and Gat (16%) systems.

## DISTRIBUTION OF PTS PERMEASE TYPES IN VARIOUS BACTERIAL KINGDOMS

The occurrence of these permease types was analyzed according to organismal type and PTS permease family (Table 5). *Actinobacteria* had about equal numbers of Glc- and Fru-type systems, with the Asc and Man families showing less but substantial representation. Only one or two members had representatives of the remaining three families. Among the *Firmicutes*, the order was Glc > Fru > Lac > Man > Asc > Gat > Gut. All three classes of *Firmicutes* (bacilli, clostridia, and *Mollicutes*) had more glucose-type systems than any other type, but in the order *Bacillales*, Lac systems were more prevalent than Fru systems, which were much more common than the Man-type systems. In contrast, the order *Lactobacillales*, as well as the clostridial class, had far more Man systems than either Lac- or Fru-type systems, which were present in about equal numbers. This observation complements and expands the suggestion of Zuniga et al. (114) that the mannose-type systems may have played a role in the establishment of symbiotic relationships between bacteria and a wide spectrum of eukaryotes. Finally, the *Mollicutes* had only Glc-, Fru-, and Asc-type systems, at decreasing frequencies in that order. *Proteobacteria* exhibited a profile quite different from those of the *Firmicutes* and *Actinobacteria*. They were found to possess far more Fru-type systems than any other type. Glc-type systems were of secondary importance, followed by Man, Asc, and Lac systems, in that order. However, the prevalence of Fru-type systems in *Proteobacteria* proved to be due solely to their high occurrence in the γ-proteobacterial subdivision, for which many fully sequenced genomes are available. The alpha and beta subdivisions had a preponderance of Glc-type systems, while the one δ-proteobacterium with a complete PTS had only a Man-type system.

## NUMBERS OF PTS PROTEIN-ENCODING GENES VERSUS GENOME SIZE AND PHYSIOLOGY

Figure 3A shows the numbers of PTS protein-encoding genes plotted versus genome size. There is no good correlation, as some of the organisms with the smallest genomes have a good representation of PTS constituents, while some organisms with large genomes have none at all. Instead, it appears

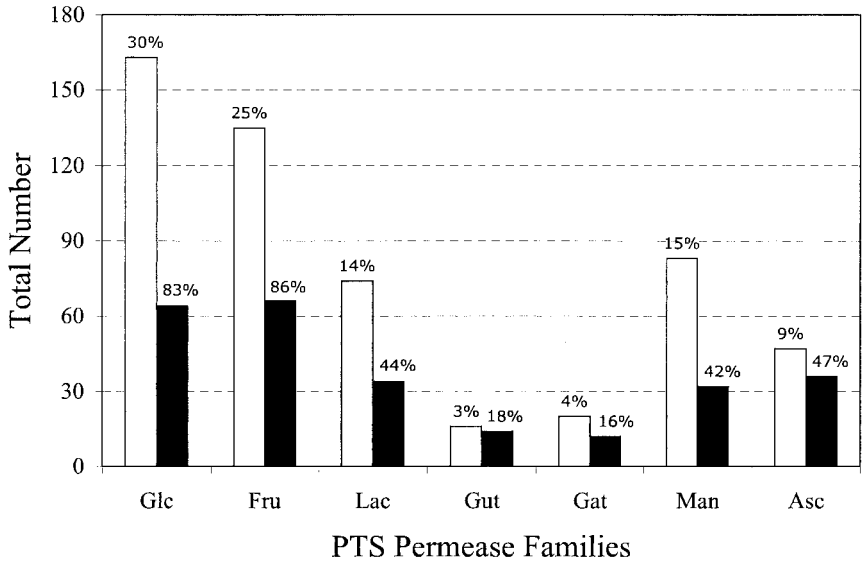



FIG. 2. Relative distribution of PTS permease families in bacterial genomes. The occurrence of the constituent permeases of the seven different PTS permease families (Glc, glucose; Fru, fructose; Lac, lactose; Gut, glucitol; Gat, galactitol; Man, mannose; Asc, ascorbate) was analyzed in the 77 bacterial species that were found to encode at least one putative complete PTS transport system. Only complete PTS permease systems were counted for this analysis; incomplete enzyme II complexes and orphan enzyme II constituents were not tabulated. Total numbers of complete PTS permeases (white bars) as well as total numbers of organisms that encode members of the different families (black bars) are presented. Values over the white bars indicate percentages of the total numbers of complete permease systems. Values over the black bars indicate percentages of the 77 organisms that have homologues of a particular family of permeases.

that the PTS protein content in a bacterium correlates best with the mode of carbohydrate metabolism utilized by that organism. Bacteria that rely on anaerobic sugar metabolism via glycolysis for energy production generally have the most PTS permeases, as discussed previously (56, 57). Thus, almost all bacteria that possess the PTS either are capable of anaerobic sugar utilization via glycolysis or have close bacterial relatives that are capable of such metabolism, while many of the bacteria that lack the PTS either are strict aerobes or lack a complete glycolytic cycle.

We tabulated the aerobic versus anaerobic metabolic capabilities of the 136 bacterial species analyzed in this report. The results showed that only 20% of the bacteria that lack genes encoding PTS proteins and 28% of the bacteria that lack PTS transport systems are capable of anaerobic growth (Fig. 3B). However, of the bacteria that possess PTS transport capabilities, 37% encoding 3 to 9 PTS proteins, 86% encoding 10 to 30 PTS proteins, and 100% encoding 31 to 93 PTS proteins are capable of anaerobic growth (Fig. 3B). This can be explained by the fact that only anaerobic glycolysis yields two molecules of PEP per molecule of hexose metabolized. One of these PEP molecules must be used for uptake of the next sugar via the PTS, while the other is required for biosynthetic purposes. The availability of excess energetic PEP molecules, the phosphoryl bonds of which have higher free energies than those of ATP, presumably provided the basis for the establishment and selective expansion of the PTS.

## ORGANISMS LACKING PTS HOMOLOGUES

All organisms whose genomes lack genes encoding identifiable PTS protein homologues are listed in Table 6. These include all archaea and eukaryotes examined, as noted above, as well as 30 bacterial species. Bacteria that totally lack PTS homologues include five actinobacteria of the genuses *Mycobacterium* (4) and *Tropheryma* (1), all six cyanobacteria examined, one *Mollicute* species (onion yellows phytoplasma 0Y-M),

TABLE 5. Numbers and types of complete PTS permeases identified in each of the bacterial kingdoms[a]

| Taxonomic group | No. of complete PTS permeases | | | | | | |
|---|---|---|---|---|---|---|---|
| | Glc | Fru | Lac | Gut | Gat | Man | Asc |
| Actinobacteria | 12 | 11 | 2 | 1 | 1 | 3 | 4 |
| *Firmicutes* | 111 | 63 | 57 | 8 | 12 | 51 | 21 |
| *Bacillales* | 37 | 29 | 36 | 4 | 9 | 11 | 7 |
| *Lactobacillales* | 46 | 17 | 17 | 4 | 3 | 33 | 6 |
| Clostridia | 9 | 5 | 4 | | | 7 | 1 |
| *Mollicutes* | 19 | 12 | | | | | 7 |
| Proteobacteria | 37 | 55 | 14 | 7 | 7 | 28 | 22 |
| Alpha subdivision | 2 | | | 1 | | | |
| Beta subdivision | 5 | 3 | | | | | |
| *Burkholderiales* | 3 | 1 | | | | | |
| *Neisseriales* | 2 | 2 | | | | | |
| Gamma subdivision | 30 | 52 | 14 | 6 | 7 | 27 | 22 |
| *Enterobacteriales* | 22 | 27 | 11 | 5 | 7 | 21 | 14 |
| *Pasteurellales* | 3 | 4 | | 1 | | 4 | 4 |
| *Pseudomonadales* | 1 | 3 | | | | | |
| *Vibrionales* | 3 | 15 | 3 | | | 2 | 4 |
| *Xanthomonadales* | | 2 | | | | | |
| Delta subdivision | | | | | | 1 | |
| Spirochetes | 2 | 4 | 1 | | | | |
| Unclassified bacteria | 1 | 2 | | | | 1 | |
| Total | 163 | 135 | 74 | 16 | 20 | 83 | 47 |

[a] PTS permeases are classified according to the seven PTS permease families, as specified in the transporter classification database (TCDB) (10, 77).

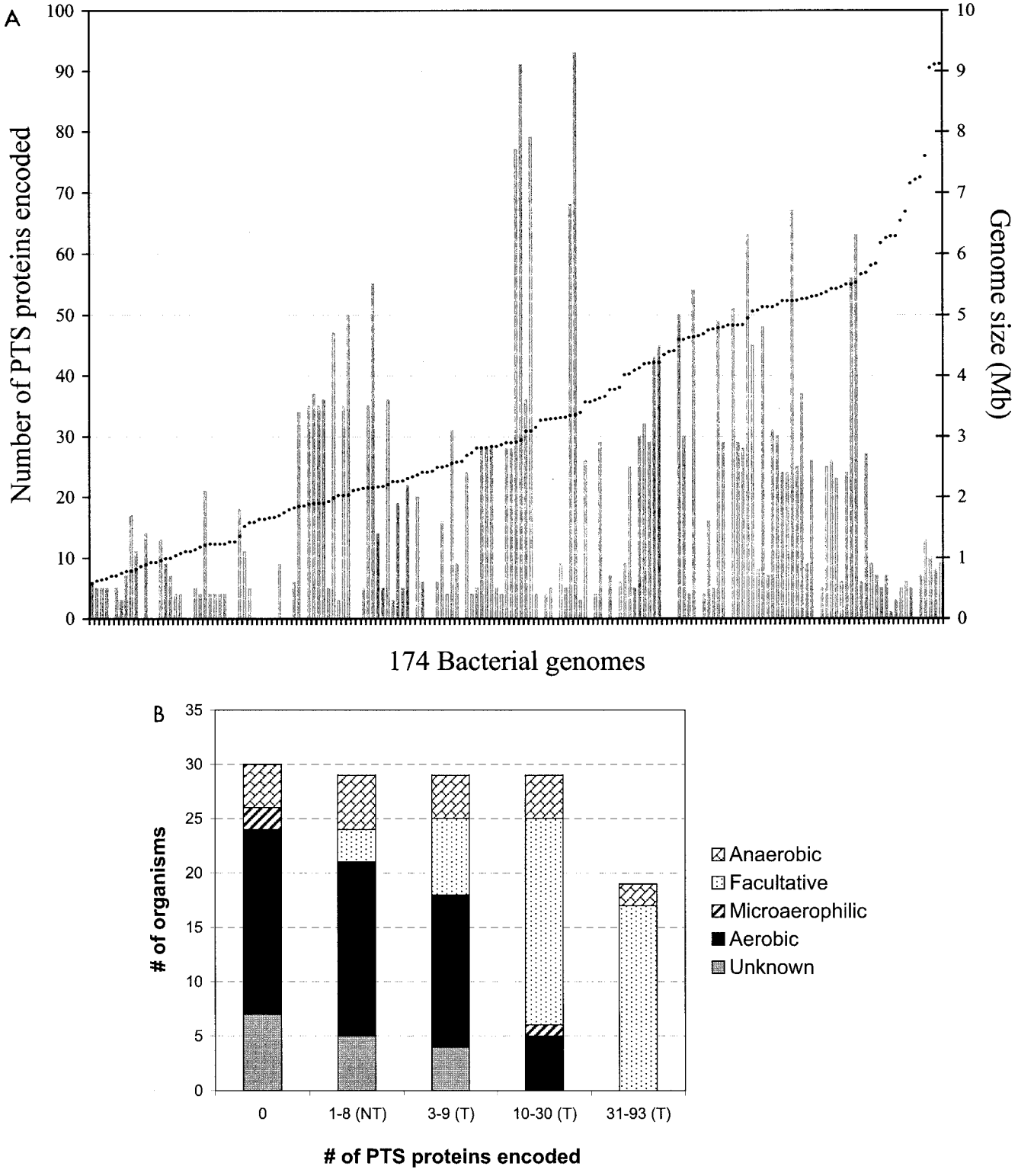



FIG. 3. (A) PTS-encoding capacities of bacterial genomes. The total number of PTS protein homologues identified in an organism (bars) and the genome size, in mega-bp (dots), are plotted. Each point on the *x* axis indicates a different genome, with the genome size increasing from left to right, as indicated. (B) Correlation between the numbers of PTS proteins encoded in the various genomes and the oxygen requirements of the bacteria. Numbers along the *x* axis indicate the numbers of PTS proteins encoded in the genomes. "NT" indicates that the bacteria do not possess PTS permeases, while "T" indicates the presence of at least one PTS transport system. Bars are shaded according to metabolic capability, such as an aerobic, anaerobic, microaerophilic, facultative, or unknown oxygen requirement.

TABLE 6. Organisms lacking PTS homologues

| Taxonomic group and organism | Genome size (Mb) |
|---|---|
| Bacteria | |
| Actinobacteria | |
| *Mycobacterium avium* subsp. *paratuberculosis* k10 | 4.83 |
| *Mycobacterium bovis* AF2122/97 | 4.35 |
| *Mycobacterium leprae* TN | 3.27 |
| *Mycobacterium tuberculosis* H37Rv | 4.41 |
| *Tropheryma whipplei* Twist | 0.93 |
| Cyanobacteria | |
| *Gloeobacter violaceus* PCC 7421 | 4.66 |
| *Nostoc* sp. strain PCC 7120 | 7.21 |
| *Prochlorococcus marinus* MIT 9313 | 2.41 |
| *Synechococcus* sp. strain WH 8102 | 2.43 |
| *Synechocystis* sp. strain PCC 6803 | 3.57 |
| *Thermosynechococcus elongatus* BP-1 | 2.59 |
| *Firmicutes*, *Onion yellows phytoplasma* OY-M | 0.86 |
| Proteobacteria | |
| Alpha subdivision | |
| *Anaplasma marginale* St. Maries | 1.20 |
| *Rickettsia conorii* Malish 7 | 1.27 |
| *Rickettsia prowazekii* Madrid E | 1.11 |
| *Rickettsia typhi* Wilmington | 1.11 |
| *Wolbachia* (endosymbiont of *Drosophila melanogaster*) | 1.27 |
| Gamma subdivision | |
| *Wigglesworthia glossinidia* (endosymbiont of *Glossina* [tsetse fly]) | 0.70 |
| *Methylococcus capsulatus* Bath | 3.30 |
| Delta subdivision | |
| *Bdellovibrio bacteriovorus* HD100 | 3.78 |
| *Desulfotalea psychrophila* LSv54 | 3.66 |
| Epsilon subdivision | |
| *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 | 1.64 |
| *Helicobacter hepaticus* ATCC 51449 | 1.80 |
| *Helicobacter pylori* 26695 | 1.67 |
| *Wolinella succinogenes* DSM 1740 | 2.11 |
| Unclassified bacteria | |
| *Aquifex aeolicus* VF5 | 1.59 |
| *Bacteroides fragilis* YCH46 | 5.31 |
| *Porphyromonas gingivalis* W83 | 2.34 |
| *Thermotoga maritima* MSB8 | 1.86 |
| *Thermus thermophilus* HB27 | 2.13 |
| Archaea | |
| *Aeropyrum pernix* K1 | 1.67 |
| *Archaeoglobus fulgidus* DSM 4304 | 2.18 |
| *Halobacterium* sp. strain NRC-1 | 2.57 |
| *Methanocaldococcus jannaschii* DSM 2661 | 1.74 |
| *Methanococcus maripaludis* S2 | 1.66 |
| *Methanopyrus kandleri* AV19 | 1.69 |
| *Methanosarcina acetivorans* C2A | 5.75 |
| *Methanosarcina mazei* Go1 | 4.10 |
| *Methanothermobacter thermautotrophicus* Delta H | 1.75 |
| *Nanoarchaeum equitans* Kin4-M | 0.49 |
| *Picrophilus torridus* DSM 9790 | 1.55 |
| *Pyrobaculum aerophilum* IM2 | 2.22 |
| *Pyrococcus abyssi* GE5 | 1.77 |
| *Pyrococcus furiosus* DSM 3638 | 1.91 |
| *Pyrococcus horikoshii* OT3 | 1.74 |
| *Sulfolobus solfataricus* P2 | 2.99 |
| *Sulfolobus tokodaii* 7 | 2.69 |
| *Thermoplasma acidophilum* DSM 1728 | 1.56 |
| *Thermoplasma volcanium* GSS1 | 1.58 |
| Eukaryotes | |
| *Anopheles gambiae* (insect) | 278 |
| *Arabidopsis thaliana* (plant) | 125 |
| *Caenorhabditis elegans* (nematode) | 97 |
| *Drosophila melanogaster* (insect) | 122 |
| *Encephalitozoon cuniculi* (fungus) | 2.5 |
| *Homo sapiens* (animal) | 3,150 |
| *Saccharomyces cerevisiae* (fungus) | 13 |
| *Neurospora crassa* (fungus) | 40 |
| *Plasmodium falciparum* (protozoan) | 23 |
| *Schizosaccharomyces pombe* (fungus) | 14 |

several proteobacteria of the alpha (five), gamma (two), delta (two), and epsilon (four) subdivisions, and five evolutionarily divergent bacteria of the genuses *Aquifex*, *Bacteroides*, *Porphyromonas*, *Thermatoga*, and *Thermus*. In the case of *Bacteroides thetaiotaomicron*, a single homologue of a galactitol IIC protein (GI 29349515) was found. This homologue cannot function via a PTS-dependent mechanism since this organism lacks all recognizable PTS phosphoryl transfer proteins. We have noted that the encoding gene is in a monocistronic operon with a good promoter, suggesting that this IIC homologue may function as a secondary carrier, as discussed previously (33, 81). Therefore, *B. thetaiotaomicron*, like *Bacteroides fragilis*, probably lacks PTS proteins altogether. Except for cyanobacteria, the ε-*Proteobacteria*, and several primitive bacteria in the "unclassified bacteria" category (Table 6), PTS homologues are found in all of the major bacterial kingdoms for which sufficient sequence data are available. *Methylococcus capsulatus* encodes two homologues of the Dha proteins (see below). However, it lacks all other PTS protein homologues, including EI and HPr. It is therefore unlikely that this organism has a functional phosphoryl transfer chain.

## BACTERIA WITH CYTOPLASMIC PTS PROTEIN HOMOLOGUES BUT NO RECOGNIZABLE PTS TRANSPORTERS

Twenty-nine bacteria were found to have PTS phosphoryl transfer proteins but to lack the complete complement of enzymes necessary for sugar transport (Table 7). As noted above, *Bacteroides thetaiotaomicron* possesses only a $IIC^{Gat}$ homologue and may use this protein as a secondary carrier. A second organism, *Ureaplasma parvum*, possesses only $IIB^{Glc}$ (GI 13357736), HPr (GI 13358152), and HprK (GI 13357632). In this organism, the HPr and HprK proteins may function in catabolite repression by a PTS-mediated sugar transport-independent mechanism (48), while $IIB^{Glc}$ may be a relic of a complete enzyme $II^{Glc}$, the other components of which were lost during genome minimalization (51).

Of the remaining 27 organisms presented in Table 7, all are gram-negative bacteria. They include proteobacteria, chlamydiae, spirochetes, and a green photosynthetic bacterium. All of these organisms encode within their genomes at least one enzyme I homologue (I or $I^{Ntr}$) and at least one HPr homologue (HPr or NPr) (65). Three of these bacteria (*Pirellula* sp., *Geobacter sulfurreducens*, and *Bradyrhizobium japonicum*) have two EIs, and four of them (*Parachlamydia* sp., *Bartonella henselae*, *Bartonella quintana*, and *Bradyrhizobium japonicum*) have two HPrs. Furthermore, all but two groups of these bacteria, the chlamydial group (except for *Parachlamydia*) and *Pirellula* sp., have HprK. Finally, these organisms may possess zero to three IIA proteins ($IIA^{Ntr}$ and/or $IIA^{Fru}$ and/or $IIA^{Man}$). Only two of these bacteria (*Coxiella burnetii* and *Legionella pneumophila*) lack an identifiable IIA protein. In *Agrobacterium tumefaciens*, the $EI^{Ntr}$ protein (GI 17937868) is encoded on the linear chromosome, while the rest of the PTS proteins are encoded on the circular chromosome.

Of the IIA homologues, *Xylella* species have only a single $IIA^{Man}$ protein while *Leptospira* species have only a single $IIA^{Ntr}$ protein. All others have either two $IIA^{Ntr}$ proteins, two $IIA^{Fru}$ proteins, one each of the $IIA^{Ntr}$ and $IIA^{Fru}$ proteins,

TABLE 7. Organisms lacking PTS permeases but possessing PTS phosphoryl transfer proteins

| Organism | Protein(s) present[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| | EI and/or HPr | $EI^{Ntr}$, Npr, and/or $IIA^{Ntr}$ | Glc | Fru | Gat | Man | HprK |
| Chlamydiae | | | | | | | |
| *Chlamydia muridarum* | Both | $IIA^{Ntr}$ (2) | | | | | |
| *Chlamydia trachomatis* D/UW-3/CX | Both | $IIA^{Ntr}$ | | IIA | | | |
| *Chlamydophila caviae* GPIC | Both | $IIA^{Ntr}$ | | IIA | | | |
| *Chlamydophila pneumoniae* AR39 | Both | $IIA^{Ntr}$ | | IIA | | | |
| *Parachlamydia* sp. strain UWE25 | Both, HPr | $IIA^{Ntr}$ (2) | | | | | HprK |
| *Firmicutes* organism *Ureaplasma parvum* serovar 3 strain ATCC 700970 | HPr | | IIB | | | | HprK |
| Proteobacteria | | | | | | | |
| Alpha subdivision | | | | | | | |
| *Agrobacterium tumefaciens* C58 | | All 3 | | | | IIA | HprK |
| *Bartonella henselae* Houston-1 | Both | NPr, $IIA^{Ntr}$ | | IIA | | IIA | HprK |
| *Bartonella quintana* Toulouse | Both | NPr, $IIA^{Ntr}$ | | IIA | | IIA | HprK |
| *Bradyrhizobium japonicum* USDA 110[b] | Both | All 3 | | | | IIA | HprK |
| *Brucella melitensis* 16M | | All 3 | | | | IIA | HprK |
| *Brucella suis* 1330 | | All 3 | | | | IIA | HprK |
| *Rhodopseudomonas palustris* CGA009 | | All 3, $IIA^{Ntr}$ | | | | IIA | HprK |
| *Sinorhizobium meliloti* 1021 | | All 3 | | | | IIA | HprK (2) |
| Beta subdivision | | | | | | | |
| *Bordetella bronchiseptica* RB50 | EI | NPr, $IIA^{Ntr}$ | | | | IIA | HprK |
| *Bordetella parapertussis* 12822 | EI | NPr, $IIA^{Ntr}$ | | | | IIA | HprK |
| *Bordetella pertussis* Tohama I | EI | NPr, $IIA^{Ntr}$ | | | | IIA | HprK |
| *Neisseria meningitidis* MC58 | EI | NPr, $IIA^{Ntr}$ | | | | IIA | HprK |
| *Nitrosomonas europaea* ATCC 19718 | EI | NPr, $IIA^{Ntr}$ | | | | IIA | HprK |
| Gamma subdivision | | | | | | | |
| *Coxiella burnetii* RSA 493 | HPr | $EI^{Ntr}$ | | | | | HprK |
| *Legionella pneumophila* subsp. *pneumophila* strain Philadelphia 1 | HPr | $EI^{Ntr}$ | | | | | HprK |
| *Xylella fastidiosa* 9a5c | EI | NPr | | | | IIA | HprK |
| Delta subdivision organism *Geobacter sulfurreducens* PCA | EI (2) | NPr, $IIA^{Ntr}$ | | | | IIA | HprK |
| Spirochetes | | | | | | | |
| *Leptospira interrogans* serovar *Lai* strain 56601 | EI | NPr, $IIA^{Ntr}$ | | | | | HprK |
| *Treponema denticola* ATCC 35405 | Both | $IIA^{Ntr}$ (2) | | | | | HprK |
| *Treponema pallidum* subsp. *pallidum* strain Nichols | Both | $IIA^{Ntr}$ (2) | | | | | HprK |
| Unclassified bacteria | | | | | | | |
| *Bacteroides thetaiotaomicron* VPI-5482 | | | | | IIC | | |
| *Chlorobium tepidum* TLS | Both | | | IIA (2) | | | HprK |
| *Pirellula* sp. | Both, EI | $IIA^{Ntr}$ (2) | | | | | |

[a] Numbers in parentheses indicate the numbers of proteins or systems.
[b] This organism possesses a putative complete DHA PTS.

one each of the $IIA^{Ntr}$ and $IIA^{Man}$ proteins, one $IIA^{Ntr}$, one $IIA^{Fru}$, and one $IIA^{Man}$ protein, or two $IIA^{Ntr}$ and one $IIA^{Man}$ protein (Table 7). The species with one $IIA^{Fru}$, one $IIA^{Ntr}$, and one $IIA^{Man}$ homologue are the two *Bartonella* species, while *Rhodopseudomonas palustris* has two $IIA^{Ntr}$ and one $IIA^{Man}$ protein. The fact that most organisms possess at least two homologous or nonhomologous IIA proteins suggests that the PTS phosphoryl transfer chain functions in a regulatory capacity that depends on specific functional interactions between these proteins. These putative interactions could, for example, be antagonistic or synergistic.

In summary, bacteria that possess PTS phosphoryl transfer proteins but lack PTS permeases always have at least one enzyme I (I or $I^{Ntr}$) and one HPr (HPr or NPr). They usually have two IIA proteins and one HprK. Only for one such organism, *Treponema denticola*, have the catalytic activities of the phosphoryl transfer proteins been demonstrated (23). *Treponema pallidum* lacks an active enzyme I homologue and instead possesses a *ptsI* pseudogene (23). We suggest that in most cases, these PTS phosphoryl transfer proteins act together, comprising a single unified phosphoryl transfer chain (63, 65) that acts biochemically as shown in Fig. 1B. Among these bacteria, two distinct, independently functioning phosphoryl transfer chains (65) appear to be present only in *Bradyrhizobium japonicum*. These two chains may catalyze phosphoryl transfer as shown in Fig. 1C. It is also interesting that among these bacteria, *B. japonicum* is the only one that possesses a complete putative DHA PTS (see below). We suggest that pathway 1 functions to phosphorylate DHA while pathway 2 functions to coordinate carbon and nitrogen metabolism (46, 47, 63, 65).

TABLE 8. Bacteria with just one or two types of PTS enzyme II complexes

| Organism | Protein(s) present[a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | EI and/or HPr | $EI^{Ntr}$, NPr, and/or $IIA^{Ntr}$ | Glc | Fru | Lac | Gut | Man | Asc | HprK |
| Bacteria with just one type of PTS permease | | | | | | | | | |
| *Acinetobacter* sp. strain ADP1 | Both | $EI^{Ntr}$, NPr | | 1 | | | | | |
| *Bifidobacterium longum* NCC2705 | Both | | 1 | | | | | | |
| *Burkholderia mallei* ATCC 23344 | Both, EI | NPr, $IIA^{Ntr}$ | 1 | | | | A | | HprK |
| *Burkholderia pseudomallei* K96243 | Both, EI | NPr, $IIA^{Ntr}$ | 1 | | | | A | | HprK |
| "*Candidatus* Blochmannia floridanus" | Both | | | | | | 1 | | |
| *Caulobacter crescentus* CB15 | Both (2) | All 3 | 2 | | | | A | | HprK |
| *Clostridium tetani* E88 | Both | | 1 | | | | | | HprK |
| *Deinococcus radiodurans* R1 | Both | | | 1 | | | | | HprK |
| *Desulfovibrio vulgaris* subsp. *vulgaris* strain Hildenborough | Both (2) | $IIA^{Ntr}$ | | | | | 1 | | |
| *Haemophilus influenzae* Rd KW20 | Both, FPr | $IIA^{Ntr}$ | A | 1 | | | | | |
| *Leifsonia xyli* subsp. *xyli* CTCB07 | Both, HPr | | | 2 | B | | | AB | |
| *Mesorhizobium loti* MAFF303099 | Both, EI | All 3 | | | | 1, C | A | | HprK (2) |
| *Nocardia farcinica* | Both | | | 1 | | | | | |
| *Pseudomonas putida* KT2440 | Both | All 3 | | 1 | | | | | |
| *Pseudomonas syringae* pv. *tomato* strain DC3000 | Both | All 3 | | 1 | | | | | |
| *Shewanella oneidensis* MR-1 | Both | All 3 | 1 | | | | | | |
| *Xanthomonas axonopodis* pv. *citri* strain 306 | Both, EI | NPr, $IIA^{Ntr}$ | | 1 | | | A | | HprK (3) |
| *Xanthomonas campestris* pv. *campestris* strain ATCC 33913 | Both, EI | NPr, $IIA^{Ntr}$ | | 1 | | | A | | HprK (2) |
| Bacteria with just two types of PTS permeases | | | | | | | | | |
| *Borrelia garinii* PBi | Both, HPr | | 1 | 2 | | | | | |
| *Buchnera aphidicola* APS (*Acyrthosiphon pisum*) | Both | | 1 | 1 | | | | | |
| *Chromobacterium violaceum* ATCC 12472 | Both (4), EI | NPr, $IIA^{Ntr}$ (2) | 2 | 2 | | | A | | HprK |
| *Corynebacterium diphtheriae* NCTC 13129 | Both, HPr | | 1 | | | | | 1 | |
| *Corynebacterium efficiens* YS-314 | Both | | 1 | 1 | | | | | |
| *Haemophilus ducreyi* 35000HP | Both | $IIA^{Ntr}$ | A | | | | 1 | 1 | |
| *Mesoplasma florum* L1 | Both | | 7 | 1 | | | | C | HprK |
| *Mycoplasma gallisepticum* R | Both | | 1 | 1 | | | | | HprK |
| *Mycoplasma genitalium* G-37 | Both | | 1 | 1 | | | | | HprK |
| *Mycoplasma mobile* 163K | Both | | 1 | 1 | | | | | HprK |
| *Pseudomonas aeruginosa* PAO1 | Both (2) | All 3 | 1 | 1 | | | | | |
| *Ralstonia solanacearum* GMI1000 | Both (2), EI | NPr, $IIA^{Ntr}$ | 1 | 1 | | | | C | HprK |
| *Streptomyces avermitilis* MA-4680 | Both | | 1 | 1 | | | | | HprK |

[a] Numbers in parentheses indicate the numbers of proteins or systems. Numbers in columns 4 to 9 indicate the numbers of complete PTS permeases. A, B, and C refer to IIA, IIB, and IIC orphan proteins, respectively, in addition to or without complete PTS enzyme II complexes. Organisms within each of the two subcategories are arranged alphabetically.

## BACTERIA WITH A COMPLETE PTS PHOSPHORYL TRANSFER CHAIN AND JUST ONE OR TWO TYPES OF PTS PERMEASE

Complete PTS permeases (enzyme II complexes) consist of IIA, IIB, and IIC (as well as IID in the case of the Man family) domains that may occur as separate peptides or may be fused in various combinations into a smaller number of polypeptide chains (see below). Table 8 lists the organisms that have only one or two types of complete PTS permease in addition to the PTS energy-coupling proteins. Of the 77 bacterial species that were found to encode at least one complete PTS transport system, 31 were found to encode just one or two types of PTS permease. Eighteen bacterial species possess just a single complete type of PTS permease. All of these organisms have both enzyme I and HPr, and many of the gram-negative *Proteobacteria* also have a partial or complete nitrogen regulatory phosphoryl transfer chain including $I^{Ntr}$, NPr, and $IIA^{Ntr}$. Like the case for *E. coli* (63, 65), these organisms may possess two independently functioning phosphoryl transfer chains, one for sugar transport and one for regulation (45–47). Nine species have just one or two fructose-type PTS permeases, while six possess only one or two glucose-type systems; two have just a single mannose system, and one has only one glucitol-type system. Except for *Caulobacter crescentus*, which has two Glc-type systems, and *Mesorhizobium loti*, which has a Gut-type system, all other α-*Proteobacteria* examined either possess just the regulatory PTS proteins (Table 7) or lack PTS homologues altogether (Table 6). "*Candidatus* Blochmannia floridanus," an *Enterobacteriaceae* member, is one of the organisms that has just a single Man-type PTS. It is possible that it once possessed Fru- and/or Glu-type systems since many of the *Enterobacteriaceae* analyzed encode several Fru-, Glu-, and Man-type PTS permeases (Table 9). None of the organisms analyzed possesses only a single lactose-, galactitol-, or L-ascorbate-type system. However, orphan IIA and/or IIB and/or IIC homologues of these systems were often found. These may be residues of genome minimalization (51) where the other constituents of complete enzyme II complexes were lost.

TABLE 9. Summary of complete PTSs in PTS-encoding organisms

| Taxonomic group and organism | Protein or system present[a] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | EI and/or HPr | $EI^{Ntr}$, NPr, and/or $IIA^{Ntr}$ | Glc | Fru | Lac | Gut | Gat | Man | Asc | HprK |
| Actinobacteria | | | | | | | | | | |
| *Bifidobacterium longum* NCC2705 | Y | | 1 | | | | | | | |
| *Corynebacterium diphtheriae* NCTC 13129 | Y | | 1 | | | | | | 1 | |
| *Corynebacterium efficiens* YS-314 | Y | | 1 | 1 | | | | | | |
| *Corynebacterium glutamicum* ATCC 13032 | Y | | 2 | 1 | | | | | 1 | |
| *Leifsonia xyli* subsp. *xyli* CTCB07 | Y | | | 2 | | | | | | |
| *Nocardia farcinica* | Y | | | 1 | | | | | | |
| *Propionibacterium acnes* KPA171202 | Y | | 3 | 3 | 1 | 1 | 1 | | 1 | |
| *Streptomyces avermitilis* MA-4680 | Y | | 1 | 1 | | | | | | Y |
| *Streptomyces coelicolor* A3(2) | Y | | 1 | 1 | | | | | 1 | |
| *Symbiobacterium thermophilum* IAM 14863 | Y | | 2 | 1 | 1 | | | 3 | | |
| *Firmicutes* | | | | | | | | | | |
| *Bacillales* | | | | | | | | | | |
| *Bacillus anthracis* Ames | Y | | 2 | 1 | 3 | | | | | Y |
| *Bacillus cereus* ATCC 10987 | Y | | 2 | 1 | 3 | | | | 1 | Y |
| *Bacillus halodurans* C-125 | Y | | 4 | 2 | 2 | 1 | 1 | | 1 | Y |
| *Bacillus licheniformis* DSM 13 | Y | | 5 | 4 | 3 | 1 | | 1 | | Y |
| *Bacillus subtilis* subsp. *subtilis* 168 | Y | | 5 | 3 | 2 | | | 1 | | Y |
| *Bacillus thuringiensis* serovar konkukian strain 97-27 | Y | | 3 | 1 | 3 | | | | | Y |
| *Listeria innocua* Clip11262 | Y | | 2 | 4 | 8 | 1 | 3 | 4 | 1 | Y |
| *Listeria monocytogenes* EGD-e | Y | | 6 | 6 | 8 | 1 | 3 | 4 | 2 | Y |
| *Oceanobacillus iheyensis* HTE831 | Y | | 2 | 2 | 2 | | 1 | 1 | 1 | Y |
| *Staphylococcus aureus* subsp. *aureus* MRSA252 | Y | | 4 | 3 | 1 | | 1 | | 1 | Y |
| *Staphylococcus epidermis* ATCC 12228 | Y | | 2 | 2 | 1 | | | | | Y |
| *Lactobacillales* | | | | | | | | | | |
| *Enterococcus faecalis* V583 | Y | | 6 | 4 | 1 | 1 | | 12 | 1 | Y |
| *Lactobacillus johnsonii* MCC 533 | Y | | 7 | 2 | 1 | | 1 | 5 | | Y |
| *Lactobacillus plantarum* WCFS1 | Y | | 14 | 2 | 2 | 2 | 2 | 2 | 1 | Y |
| *Lactococcus lactis* subsp. *lactis* III403 | Y | | 3 | 1 | 1 | | | 1 | | Y |
| *Streptococcus agalactiae* NEM316 | Y | | 4 | 1 | 2 | | | 3 | 1 | Y |
| *Streptococcus mutans* UA159 | Y | | 4 | 3 | 2 | 1 | | 3 | 1 | Y |
| *Streptococcus pneumoniae* TIGR4 | Y | | 4 | 3 | 5 | | | 4 | 1 | Y |
| *Streptococcus pyogenes* MGAS315 | Y | | 4 | 1 | 3 | | | 3 | 1 | Y |
| Clostridia | | | | | | | | | | |
| *Clostridium acetobutylicum* ATCC 824 | Y | | 6 | 2 | 2 | | | 2 | | Y |
| *Clostridium perfringens* 13 | Y | | 1 | 1 | | | | 4 | 1 | Y |
| *Clostridium tetani* E88 | Y | | 1 | | | | | | | Y |
| *Thermoanaerobacter tengcongensis*[b] | Y | | 1 | 2 | 2 | | | 1 | | Y |
| *Mollicutes* | | | | | | | | | | |
| *Mesoplasma florum* L1 | Y | | 7 | 1 | | | | | | Y |
| *Mycoplasma gallisepticum* R | Y | | 1 | 1 | | | | | | Y |
| *Mycoplasma genitalium* G-37 | Y | | 1 | 1 | | | | | | Y |
| *Mycoplasma hyopneumoniae* 232 | Y | | 1 | 2 | | | | | 2 | |
| *Mycoplasma mobile* 163K | Y | | 1 | 1 | | | | | | Y |
| *Mycoplasma mycoides* subsp. *mycoides* SC strain PG1 | Y | | 1 | 2 | | | | | 1 | Y |
| *Mycoplasma penetrans* HF-2 | Y | | 5 | 2 | | | | | 2 | Y |
| *Mycoplasma pneumoniae* M129 | Y | | 1 | 1 | | | | | 1 | Y |
| *Mycoplasma pulmonis* UAB CTIP | Y | | 1 | 1 | | | | | 1 | Y |
| Proteobacteria | | | | | | | | | | |
| Alpha subdivision | | | | | | | | | | |
| *Caulobacter crescentus* CB15 | Y | | 2 | | | | | | | Y |
| *Mesorhizobium loti* MAFF 303099 | Y | Y | | | | 1 | | | | Y |
| Beta subdivision | | | | | | | | | | |
| *Burkholderiales* | | | | | | | | | | |
| *Burkholderia mallei* ATCC 23344 | Y | | 1 | | | | | | | Y |
| *Burkholderia pseudomallei* K96243 | Y | | 1 | | | | | | | Y |
| *Ralstonia solanacearum* GMI1000 | Y | | 1 | 1 | | | | | | Y |
| *Neisseriales* | | | | | | | | | | |
| *Chromobacterium violaceum* ATCC 12472 | Y | | 2 | 2 | | | | | | Y |
| Gamma subdivision | | | | | | | | | | |
| *Enterobacteriales* | | | | | | | | | | |
| *Buchnera aphidicola* APS (*Acyrthosiphon pisum*) | Y | | 1 | 1 | | | | | | |
| "*Candidatus* Blochmannia floridanus" | Y | | | | | | | 1 | | |

*Continued on following page*

TABLE 9—*Continued*

| Taxonomic group and organism | Protein or system present[a] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | EI and/or HPr | $EI^{Ntr}$, NPr, and/or $IIA^{Ntr}$ | Glc | Fru | Lac | Gut | Gat | Man | Asc | HprK |
| *Erwinia carotovora* subsp. *atroseptica* SCRI1043 | Y | Y | 5 | 3 | 3 | | | 1 | 2 | |
| *Escherichia coli* CFT073 | Y | Y | 3 | 5 | 3 | 1 | 2 | 4 | 2 | |
| *Photorhabdus luminescens* subsp. *laumondii* TTO1 | Y | | 2 | 1 | 1 | | | 2 | 1 | |
| *Salmonella enterica* subsp. *enterica* serovar Typhi Ty2 | Y | Y | 2 | 3 | 1 | 1 | 2 | 2 | 3 | |
| *Salmonella enterica* serovar Typhimurium LT2 | Y | Y | 2 | 4 | 1 | 2 | 2 | 4 | 3 | |
| *Shigella flexneri* 2a strain 301 | Y | Y | 3 | 4 | 1 | 1 | 1 | 3 | 1 | |
| *Yersinia pestis* KIM | Y | Y | 2 | 3 | 1 | | | 2 | 1 | |
| *Yersinia pseudotuberculosis* IP 32953 | Y | Y | 2 | 3 | | | | 2 | 1 | |
| *Pasteurellales* | | | | | | | | | | |
| *Haemophilus ducreyi* 35000HP | Y | | | | | | | 1 | 1 | |
| *Haemophilus influenzae* Rd KW20 | Y | | | 1 | | | | | | |
| *Mannheimia succiniciproducens* MBEL55E | Y | | 2 | 2 | | | | 2 | 2 | |
| *Pasteurella multocida* subsp. *multocida* Pm70 | Y | | 1 | 1 | | 1 | | 1 | 1 | |
| *Pseudomonadales* | | | | | | | | | | |
| *Pseudomonas aeruginosa* PAO1 | Y | Y | 1 | 1 | | | | | | |
| *Pseudomonas putida* KT2440 | Y | Y | | 1 | | | | | | |
| *Pseudomonas syringae* pv. tomato strain DC3000 | Y | Y | | 1 | | | | | | |
| *Vibrionales* | | | | | | | | | | |
| *Vibrio cholerae* O1 biovar eltor strain N16961 | Y | Y | 1 | 4 | 1 | | | | 1 | |
| *Vibrio parahaemolyticus* RIMD 2210633 | Y | Y | 1 | 5 | 1 | | | | 1 | |
| *Vibrio vulnificus* YJ016 | Y | Y | 1 | 6 | 1 | | | 2 | 2 | |
| *Xanthomonadales* | | | | | | | | | | |
| *Xanthomonas axonopodis* pv. *citri* strain 306 | Y | | | 1 | | | | | | Y |
| *Xanthomonas campestris* pv. *campestris* strain ATCC 33913 | Y | | | 1 | | | | | | Y |
| Others | | | | | | | | | | |
| *Acinetobacter* sp. strain ADP1 | Y | | | 1 | | | | | | |
| *Shewanella oneidensis* MR-1 | Y | Y | 1 | | | | | | | |
| Delta subdivision, *Desulfovibrio vulgaris* subsp. *vulgaris* Hildenborough | Y | | | | | | | 1 | | |
| *Spirochaetales* | | | | | | | | | | |
| *Borrelia burgdorferi* B31 | Y | | 1 | 2 | 1 | | | | | |
| *Borrelia garinii* PBi | Y | | 1 | 2 | | | | | | |
| Unclassified bacteria | | | | | | | | | | |
| *Deinococcus radiodurans* R1 | Y | | | 1 | | | | | | Y |
| *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586 | Y | | 1 | 1 | | | | 1 | | Y |

[a] The organisms that encode homologues of PTS proteins are listed alphabetically and grouped according to their taxonomic groups. Only one strain is listed for organisms for which multiple strains have been analyzed. "Y" indicates the presence of both EI and HPr in column 2, the presence of homologues of all three regulatory proteins, $EI^{Ntr}$, NPr, and $IIA^{Ntr}$, in column 3, and the presence of an HprK homologue in the last column. The presence of just one of the two proteins in column 2 or the presence of one or two of the three proteins in column 3 is not indicated in this table. The numbers of copies of only complete PTS enzyme IIs that include IIA, IIB, and IIC domains (as well as IID domains in the case of the Man family) encoded in the genome of an organism are indicated for the seven PTS families (Glc, glucose; Fru, fructose; Lac, lactose; Gut, glucitol; Gat, galactitol; Man, mannose; Asc, ascorbate) in their respective columns. However, organisms may encode incomplete enzyme II complexes which are not indicated in this table.

[b] The putative *gat* operon in *T. tengcongensis* encodes a $IIA^{Fru}$ homologue rather than a $IIA^{Gat}$ homologue upstream of the genes encoding the $IIB^{Gat}$ and $IIC^{Gat}$ homologues. Therefore, this system was not counted as an intact PTS permease. However, it must be noted that $IIA^{Fru}$ and $IIA^{Gat}$ share distant homology.

Thirteen bacterial species possess just two types of complete PTS permeases. In all cases, one of the systems is either a glucose- or fructose-type system, except in *Haemophilus ducreyi*, where the two systems identified are Man- and Asc-type systems. However, *Haemophilus influenzae* has a single Fru-type system. Further examination of other bacteria in the same phylogenetic group revealed that these organisms possess Fru- and Glc-type systems in addition to Man- and Asc-type permeases (Table 9). It is therefore possible that *H. ducreyi* may have lost the Fru- and Glc-type systems during evolution. Eleven species have both glucose and fructose systems, usually just one of each. However, *Mesoplasma florum* has seven glucose-like systems and one fructose-like system. *Borrelia garinii* has two fructose-like systems and one glucose-like system, while *Chromobacterium violaceum* has two of each. *Corynebacterium diphtheriae* has a glucose-type system and an L-ascorbate-type system, while *Haemophilus ducreyi* has a mannose-type system and an ascorbate-type system, as noted above (Table 8).

None of the organisms that have just two types of PTS permeases have lactose-, galactitol-, or glucitol-type systems. In fact, even orphan constituents of these systems are lacking. However, several additional PTS orphan proteins (IIA, IIB, or IIC), particularly mannose IIA homologues, are found in some of these organisms. It is likely that some of these possess specific regulatory functions, but no such function is currently recognized. Others probably represent either relics of complete PTS permeases that have been partially lost or orphan genes acquired through horizontal transfer. Interestingly, in *Deinococcus radiodurans*, all of the PTS homologues were found on the plasmid MP1.

## BACTERIA WITH A COMPLETE PTS PHOSPHORYL TRANSFER CHAIN AND MULTIPLE TYPES OF PTS PERMEASES

Table 9 lists the 77 bacterial species that encode all of the proteins required for PTS-dependent sugar transport. Of these, 46 were found to encode more than two types of PTS permease. In the table, these bacteria are grouped according to organismal phylogenetic division, and sometimes by subdivision. These will be discussed according to their taxonomic groups in the order in which they are presented in Table 9. It should be noted that the occurrence of multiple paralogues of a specific family or subfamily of PTS permeases implies differing substrate specificities, affinities, or regulatory properties. In very few cases have these been studied in detail.

### High-G+C Gram-Positive Bacteria

Among the high-G+C gram-positive *Actinobacteria*, the *Mycobacteria* and *Tropheryma* species have no recognizable PTS protein homologues (Table 6). *Bifidobacterium longum* and *Nocardia farcinica* have a single glucose-type system and a single fructose-type system, respectively. Also, these two bacteria encode the smallest number of PTS genes (three genes) among all of the bacteria that possess a complete transport PTS. Of the three *Corynebacterium* species studied, *C. efficiens* has both Glc and Fru systems, *C. diphtheriae* has Glc and Asc systems, and *C. glutamicum* has Glc, Fru, and Asc systems. The corynebacteria represent an example where several species within a single genus have different complements of PTS proteins.

*Streptomyces avermitilis* has Glc and Fru systems, and surprisingly, of the 10 *Actinobacteria* examined, this is the only one to have HprK (GI 29826878). The HprK homologue from this bacterium clusters separately but is closest to the homologues from α-*Proteobacteria* (97). This may represent a case of horizontal gene transfer. *Streptomyces coelicolor* has an Asc system in addition to Glc and Fru systems, but it lacks HprK. Of the *Actinobacteria*, *Symbiobacterium thermophilum* and *Propionibacterium acnes* have the most PTS permeases, with 7 and 10 complete systems, respectively. Both have Glc, Fru, and Lac systems, and *S. thermophilum* has three Man systems while *P. acnes* has just one each of the Gut-, Gat-, and Asc-type systems but no Man system.

*S. thermophilum* is a thermophilic bacterium that depends on microbial commensalism. Recently, a role for the Man-type PTS transporter in the establishment of symbiotic relationships has been implicated (114). The presence of three complete Man-type systems in this organism may contribute towards its commensalic growth. *S. thermophilum* shows remarkable similarity to bacilli and clostridia (low-G+C gram-positive bacteria) and is suggested to have shared a close common ancestor with the *Bacillus*/*Clostridium* group (105). It is interesting that the PTS family representation in this bacterium also resembles the PTS protein distribution in some of the bacilli and clostridia.

*P. acnes* is found ubiquitously on human skin and resides within sebaceous follicles. It has been implicated in various diseases as an opportunistic pathogen (9). Its genome encodes a great capacity to cope with changing oxygen tension, explaining its ubiquitous presence on human skin. The occurrence of a wider variety of PTS permeases may enable efficient uptake of a broad range of substrates and may allow the organism to opportunistically adapt to a pathogenic lifestyle.

All of these high-G+C gram-positive bacteria possess various orphan proteins or "partial" PTS permeases (data not shown). The $IIBC^{Glc}$ and $IIC^{Glc}$ components may be functional, using the IIA/B proteins present in one of the complete Glc systems (101). Precedence for this has been observed repeatedly, as several Glc-type systems lack their own IIA protein and use the general IIA protein of the authentic glucose system (101). However, the same phenomenon has not been documented for the other PTS permease families.

### Low-G+C Gram-Positive Bacteria

*Firmicutes* (orders *Bacillales* and *Lactobacillales* and class *Clostridia* in Table 9) generally have the energy-coupling PTS proteins enzyme I and HPr as well as HprK, and with just two exceptions, they all have Glc, Fru, and Lac systems. The two exceptions are *Clostridium tetani*, which has only a Glc system, and *Clostridium perfringens*, which has Glc and Fru systems but not a Lac system. The three clostridial species analyzed show remarkable variation in their numbers of PTS genes, and these numbers do not correlate well with the differences in their genome sizes. *C. tetani*, with a genome size of 2.87 Mb, has 4 PTS genes, *Clostridium acetobutylicum*, with a genome size of 4.13 Mb, has 30 PTS genes, and *C. perfringens*, with an intermediate genome size of 3.09 Mb, has 36 PTS genes. The distributions of various PTS permeases in these three species are remarkably different as well (Table 9).

The occurrence of 12 complete PTSs in *C. acetobutylicum* correlates with the saccharolytic lifestyle of this soil bacterium. *C. perfringens* is commonly found in animal and human gastrointestinal tracts as a member of the normal microflora. Interestingly, the presence of multiple Man-type PTS permeases in this organism correlates with the abundance of Man-type permeases in some of the other members of the normal intestinal microflora, such as lactobacilli, *Enterococcus faecalis*, and several members of the *Enterobacteriaceae*. One of the two Man systems in *C. acetobutylicum* is encoded on the plasmid pSOL1.

Most of the other low-G+C gram-positive bacteria have additional PTS permeases, and some, such as three *Listeria* species and *Lactobacillus plantarum*, have all seven PTS enzyme II complex types. The bacteria with the most *pts* genes and PTS permeases are *L. monocytogenes*, with 91 genes and 30 complete PTS permeases plus several partial systems, and *Enterococcus faecalis*, with 93 genes and 25 complete permeases plus 14 partial systems, all of which have the IIC constituent (data not shown). Several of these may be functional, using components from the complete systems (101). The maximal percentage of genetic material devoted to the PTS is observed for *L. monocytogenes*, with 3.2% of all its genes encoding PTS proteins.

Of the small-genome *Mollicutes* organisms, all but two possess complete Glc- and Fru-type PTS permeases as well as an HprK homologue (with the exception of *Mycoplasma hyopneumoniae*), and several have an Asc system as well. The two exceptions (discussed above) are listed in Tables 6 and 7. A few reports have discussed some of the PTS proteins from the different *Mycoplasma* species (26, 30, 48, 59).

The regulation of HprK in the *Mollicutes* may be different from that in other *Firmicutes* (26). The distribution and numbers of PTS permeases in the *Mollicutes* also differ from those in other *Firmicutes*. The various species within the *Mycoplasma* genus also show differences. Most surprisingly, *Mycoplasma pulmonis* and *Mycoplasma pneumoniae* have three PTS permeases each, *Mycoplasma mycoides* has four, *M. hyopneumoniae* has five, and *M. florum* and *Mycoplasma penetrans* have eight and nine, respectively. In some of these bacteria with genome sizes of 0.79 to 1.36 Mbp, a major fraction of the genetic material (over 2%) is devoted to the PTS. Their obligate parasitic lifestyle and dependence on glycolysis as the ATP-generating pathway may explain the retention of PTS genes, even under conditions that result in extensive genome reduction (21).

## Proteobacteria

Table 9 presents the PTS protein complements in a variety of *Proteobacteria* according to their subdivision classification. Within the alpha subdivision, most either lack PTS proteins altogether (Table 6) or only possess phosphoryl transfer proteins (Table 7) that presumably function in regulation. Only two have PTS permeases. *Caulobacter crescentus* has two Glc systems, while *Mesorhizobium loti* has one Gut system. As discussed below, *Bradyrhizobium japonicum* has a complete dihydroxyacetone enzyme II complex and therefore can phosphorylate this triose via the PTS.

In the beta subdivision, all organisms possess regulatory PTS proteins (Tables 7 and 9), and a few also have PTS permeases that belong exclusively to the Glc and Fru families. The maximal number of probable complete PTS permeases is four, for *Chromobacterium violaceum*. *C. violaceum* lives in tropical and subtropical regions in soil and water but can occasionally be pathogenic in immunocompromised individuals, causing diarrhea. No other β-proteobacterium has more than two complete PTS permeases. While *Ralstonia solanacearum* has a Glc- and a Fru-type system, *Burkholderia* species have only a single Glc-type system. The Glc system in *R. solanacearum* was found encoded in the plasmid pGMI-1000MP. *Burkholderia pseudomallei* is an opportunistic pathogen in humans that is usually found in terrestrial environments, while *Burkholderia mallei* is a host-adapted pathogen in animals and is not found outside the host.

In the gamma subdivision, only the endosymbiont of the tsetse fly, *Wigglesworthia glossinidia*, and the obligate methanotroph *Methylococcus capsulatus* lack PTS homologues altogether (Table 6). *Buchnera* species, which are endosymbionts of aphids, have both a single Glc system and a single Fru system. "*Candidatus* Blochmannia floridanus," an endosymbiont of the carpenter ant, has only a mannose system. All other enterobacteria have multiple PTS permeases. It is therefore clear that these endosymbionts have retained the minimal number of PTS permeases to allow utilization of a very restricted number of sugars provided by the host organism (11).

Several *E. coli*, *Shigella*, and *Salmonella* strains (see below) have all seven types of PTS permease systems. The different strains of *E. coli* probably have between 17 and 26 functional PTS permeases. These variations, observed for various *E. coli*/*Shigella* strains, suggest that the gain and loss of genetic material encoding PTS proteins have occurred frequently and repeatedly during the evolution of single species (108). Scrutiny of Table 9 will reveal that this is a common observation for many bacterial species and genera.

Other γ-*Proteobacteria* listed in Table 9 have far fewer PTS permeases. The pseudomonads, for example, have just one or two. The three species of *Pseudomonas* studied, all with large genome sizes of over 6 Mb, vary from nonpathogenic (*P. putida*) to plant pathogenic (*P. syringae*) to opportunistically human pathogenic (*P. aeruginosa*). In *P. aeruginosa*, one PTS permease is fructose specific while the other is *N*-acetylglucosamine specific (68; I. T. Paulsen, personal communication). The *Pasteurellales* have 1 to 8 PTS permeases, vibrios have 7 to 12 PTS permeases, and the two *Xanthomonas* species studied, which are found exclusively associated with their plant hosts and are not found free in the soil, have just one fructose-type system each (17, 80, 110). Among the *Pasteurellales*, "*Mannheimia succiniciproducens*" (proposed name), which is found in bovine rumen, has the largest number of PTS permeases. The rumen is the first section of the stomach in ruminants, where feed is collected for initial digestion. The presence of several PTS transporters in ruminant bacteria perhaps represents an adaptation to a nutrient-rich, oxygen-free ecological niche. *Pasteurella multocida* is found in the mucous lining of animal intestines, in the genitalia, and in respiratory tissues. It has five PTS permeases, each belonging to a different PTS family. Each of the five PTS permeases is perhaps expressed differentially depending on the host organ. *H. ducreyi* and *H. influenzae* are an obligate pathogen and an obligate parasite, respectively, in humans. The smaller numbers of PTS permeases in these bacteria may have resulted from their evolutionary adaptation to the homeostatic environment of the host.

Vibrios are abundant in marine and freshwater environments. The three species analyzed are pathogenic to humans. *Vibrio vulnificus*, which is an opportunistic pathogen causing a variety of diseases in immunocompromised patients, has Man-type PTS permeases that are lacking in the other two species. Interestingly, it also has the largest number of PTS permeases among the three *Vibrio* species analyzed. *Vibrio parahaemolyticus*, which causes gastroenteritis, and *Vibrio cholerae*, the causative agent of cholera, have similar complements of PTS proteins. The Asc systems, a few of the Fru PTS permeases, one of the two Man systems in *V. vulnificus*, and a few other orphan PTS proteins are encoded in the smaller chromosome, chromosome II, in all three species. All other PTS permeases are encoded in chromosome I.

## Spirochetes

Among the spirochetes, *Leptospira* and *Treponema* species lack PTS permeases (as discussed earlier), but *Borrelia* species have several. *Leptospira interrogans* is an obligate aerobe that is adapted for mammalian reservoir hosts. *T. denticola* is an obligate anaerobe that is commonly found in the oral cavity in humans, while *T. pallidum* is an anaerobic human pathogen that causes syphilis. *Borrelia garinii* has one Glc and two Fru systems, while *Borrelia burgdorferi* has these systems plus a Lac

TABLE 10. Organisms that encode a putative complete DHA PTS[a]

| Category | Organism | Abbreviation[b] | Dha proteins[c] | DhaM domain(s)[d] |
|---|---|---|---|---|
| A | *Bradyrhizobium japonicum* USDA 110 | Bja | All 3 | $IIA^{Dha}$ |
| | *Clostridium tetani* E88 | Cte | All 3 | $IIA^{Dha}$ |
| | *Deinococcus radiodurans* R1 | Dra | All 3 (DhaL is split) | $IIA^{Dha}$ |
| | *Enterococcus faecalis* V583 | Efa | All 3 | $IIA^{Dha}$ |
| | *Mannheimia succiniciproducens* MBEL55E | Msu | All 3 | $IIA^{Dha}$ |
| | *Mycoplasma mycoides* subsp. *mycoides* SC strain PG1 | Mmy | All 3 | $IIA^{Dha}$ |
| | *Mycoplasma penetrans* HF-2 | Mpe | All 3 | $IIA^{Dha}$ |
| | *Oceanobacillus iheyensis* HTE831 | Oih | All 3 | $IIA^{Dha}$ |
| | *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586 | Fnu | All 3 | $IIA^{Dha}$ |
| | *Staphylococcus aureus* subsp. *aureus* MRSA252 | Sau | All 3 | $IIA^{Dha}$ |
| | *Staphylococcus epidermidis* ATCC 12228 | Sep | All 3 | $IIA^{Dha}$ |
| | *Streptomyces avermitilis* MA-4680 | Sav | All 3 | $IIA^{Dha}$ |
| | *Thermoanaerobacter tengcongensis* | Tte | All 3 | $IIA^{Dha}$ |
| B | *Corynebacterium diphtheriae* NCTC 13129 | Cdi | All 3 | $IIA^{Dha}$-HPr |
| | *Leifsonia xyli* subsp. *xyli* CTCB07 | Lxy | All 3 | $IIA^{Dha}$-HPr |
| | *Desulfovibrio vulgaris* subsp. *vulgaris* Hildenborough | Dvu | All 3 | $IIA^{Dha}$-HPr-EI |
| | *Escherichia coli* CFT073 | Eco | All 3 | $IIA^{Dha}$-HPr-EIΔ |
| | *Shigella flexneri* 2a strain 2457T | Sfl | All 3 | $IIA^{Dha}$-HPr-EIΔ |
| C | *Listeria innocua* Clip11262 | Lin | All 3 (2) | $IIA^{Dha}$ |
| | *Listeria monocytogenes* EGD-e | Lmo | All 3 (2) | $IIA^{Dha}$ |
| D | *Bacillus halodurans* C-125 | Bha | All 3, DhaK | $IIA^{Dha}$ |
| | *Lactobacillus plantarum* WCFS1 | Lpl | All 3, DhaK | $IIA^{Dha}$ |
| | *Lactococcus lactis* subsp. *lactis* II1403 | Lla | All 3, DhaK | $IIA^{Dha}$ |
| | *Streptococcus agalactiae* NEM316 | Sag | All 3, DhaK | $IIA^{Dha}$ |
| | *Streptococcus pyogenes* MGAS10394 | Spy | All 3, DhaK | $IIA^{Dha}$ |
| E | *Propionibacterium acnes* KPA171202 | Pac | All 3, DhaKL | $IIA^{Dha}$ |
| | *Streptomyces coelicolor* A3 (2) | Sco | All 3, DhaKL | $IIA^{Dha}$ |
| F | *Mesorhizobium loti* MAFF303099 | Mlo | All 3, DhaK (2), DhaL, DhaKL | $IIA^{Dha}$ |

[a] Organisms encoding at least one putative complete DHA PTS are grouped into six categories (A to F) based on the presence of additional complete or incomplete putative DHA PTSs or the presence of a putative ATP-dependent system.
[b] Three-letter abbreviations used in phylogenetic analyses.
[c] The occurrence of all three (DhaK, DhaL, and DhaM) proteins, believed to be required for a PTS-dependent DHA system, as well as additional proteins, is indicated. The number of paralogous proteins encoded in the genome of an organism is indicated in parentheses.
[d] The domains within the DhaM protein are indicated. The EIΔ domain represents a C-terminally truncated PTS enzyme I domain.

system. *Borrelia* species are microaerophilic organisms that live in the gastrointestinal tract of ticks and are capable of infecting multiple hosts via tick bites. *B. burgdorferi* carries dozens of plasmids, while *B. garinii* harbors only two. Interestingly, the Lac system as well as a $IICB^{Glc}$ protein (GI 11497021) in *B. burgdorferi* is encoded in the plasmid cp26.

Another spirochete species, *Spirochaeta aurantia*, whose genome has not yet been sequenced, has only a Fru-type system with specificity for mannitol (82). In this unusual system, the synthesis of all PTS enzymes (enzyme I, HPr, and the mannitol enzyme II) as well as that of mannitol-1-P dehydrogenase is induced 200-fold by the inclusion of mannitol in the growth medium (83).

## OCCURRENCE OF DIHYDROXYACETONE PTS ENZYME II COMPLEXES

Many bacteria possess nontransporting DHA-specific enzyme II complexes. Only for *E. coli* has such a system been characterized (22, 25, 94). The system consists of three proteins, named DhaK ($IIC^{Dha}$; dihydroxyacetone binding), DhaL ($IIB^{Dha}$; ADP binding), and DhaM ($IIA^{Dha}$; phosphoryl donor for ADP phosphorylation in DhaL) (4, 22, 81). The DhaM protein of *E. coli* is a three-domain protein consisting of an N-terminal $IIA^{Dha}$, a central HPr-like domain, and a C-terminal truncated enzyme I-like domain (25). DhaL subunits exhibit sequence characteristics that distinguish them from the homologous ATP-dependent DHA kinases (3, 4, 93). The DhaK subunit covalently binds the substrate dihydroxyacetone to a histidyl residue in the protein (22). These sequence characteristics allowed us to distinguish ATP- from PEP-dependent systems and therefore to tentatively conclude that all PTS DHA enzyme II complexes consist of split DhaK/DhaL systems, while all ATP-dependent kinases have these two functional domains fused in a single polypeptide chain (DhaKL).

Table 10 presents all bacteria for which we found homologues of the subunits of the DHA PTS enzyme II complex. No such homologues were found in cyanobacteria, chlamydia, and the epsilon division of the *Proteobacteria*. Twenty-seven organisms appear to have a complete DHA PTS, and of these, 15 (including two *Mycoplasma* species) are low-G+C gram-positive bacteria. Other organisms with complete DHA PTS enzyme II complexes include (i) α-, γ-, and δ-*Proteobacteria* (six organisms), including *Bradyrhizobium japonicum* (γ), *Mannheimia succiniciproducens* (γ), *Desulfovibrio vulgaris* (δ), *E. coli* (γ), *Shigella flexneri* (γ), and *Mesorhizobium loti* (α); (ii) high-G+C gram-positive bacteria (five organisms), including *Streptomyces avermitilis*, *S. coelicolor*, *Corynebacterium diphtheriae*, *Leifsonia xyli*, and *Propionibacterium acnes*; and (iii) other divergent organisms, including *Fusobacterium nucleatum* and *Deinococcus radiodurans*. Unlike all other DhaL proteins examined, that in *D. radiodurans* is split into two polypeptide chains. These chains could only be identified using nucleotide BLAST searches because these protein fragments had not

TABLE 11. Homologues of Dha proteins in organisms that lack a complete DHA PTS[a]

| Category | Organism | Abbreviation[b] | Dha protein(s) | Presence of $IIA^{Man}$ |
|---|---|---|---|---|
| A | *Bartonella quintana* | Bqu | DhaM ($IIA^{Dha}$) | |
| | *Streptococcus pneumoniae* TIGR4 | Spn | DhaM ($IIA^{Dha}$) | |
| B | *Mycoplasma gallisepticum* R | Mga | DhaL | |
| | *Pseudomonas syringae* pv. *tomato* DC3000 | Psy | ΔDhaL[c] | |
| C | *Agrobacterium tumefaciens* C58 | Atu | DhaKL | Yes |
| | *Clostridium perfringens* 13 | Cpe | DhaKL | Yes |
| | *Erwinia carotovora* subsp. *atroseptica* SCRI1043 | Eca | DhaKL | |
| D | *Brucella melitensis* 16M | Bme | DhaK, DhaL (2) | Yes |
| | *Methylococcus capsulatus* Bath | Mca | DhaK, DhaL | |
| | *Pasteurella multocida* Pm70 | Pmu | DhaK, DhaL | |
| | *Treponema denticola* ATCC 35405 | Tde | DhaK, DhaL | |
| | *Vibrio parahaemolyticus* RIMD 2210633 | Vpa | DhaK, DhaL | |
| | *Yersinia pestis* KIM | Ype | DhaK, DhaL | |
| | *Yersinia pseudotuberculosis* IP 32953 | Yps | DhaK, DhaL | |
| E | *Bacillus cereus* ZK | Bce | DhaK, ΔDhaL, DhaKL | |
| | *Burkholderia mallei* ATCC 23344 | Bma | DhaK, DhaL, DhaKL | Yes |
| | *Burkholderia pseudomallei* K96243 | Bps | DhaK, DhaL, DhaKL | Yes |
| | *Sinorhizobium meliloti* 1021 | Sme | DhaK, DhaL (2), DhaKL | Yes |
| F | *Bacillus anthracis* A2012 | Ban | DhaK, DhaKL | |
| | *Bacillus cereus* ATCC 10987 | Bce | DhaK, DhaKL | |
| | *Bacillus thuringiensis* serovar *konkukian* strain 97-27 | Bth | DhaK, DhaKL | |

[a] Organisms are grouped into six categories (A to F) based on the type of Dha proteins present. The various Dha homologues encoded by the different organisms are listed in column 3, and the number of paralogues of a particular protein is indicated in parentheses. The presence of a separately encoded PTS $IIA^{Man}$ domain is indicated in the last column.

[b] Three-letter abbreviations used in phylogenetic analyses.

[c] *P. syringae* has an N-terminally deleted DhaL and therefore was not included in the phylogenetic analyses presented in Fig. 4.

been assigned protein accession numbers. All four putative *dha* genes in *D. radiodurans* (including the genes for DhaM and DhaK and the two split genes that together comprise DhaL) were adjacent on a plasmid, the MPI plasmid (41, 109). Splicing of the putative DhaL gene could have resulted from a sequencing error. Furthermore, since the functionality of its gene product has not been demonstrated, the significance of this observation has yet to be determined.

Category A organisms possess one and only one DHA PTS. These bacteria include two γ-*Proteobacteria*, *Fusobacterium nucleatum*, one high-G+C gram-positive bacterium, and eight low-G+C gram-positive bacteria (Table 10). DhaM in these systems consists of only a $IIA^{Dha}$ domain.

Category B organisms have just one DHA PTS, but their DhaM proteins have the $IIA^{Dha}$ domain fused to other PTS protein domains. These fused proteins include $IIA^{Dha}$-HPr fusions, found in *C. diphtheriae* (GI 38234870) and *L. xyli* (GI 50954678), a single full-length $IIA^{Dha}$-HPr-EI fusion (GI 46579394), present in *D. vulgaris*, and tridomain proteins with a truncated enzyme I, $IIA^{Dha}$-HPr-EIΔ, characteristic of the *E. coli* DhaM protein, found in γ-*Proteobacteria* (Table 10). Only category B systems have the $IIA^{Dha}$ domain fused to other PTS protein domains.

*Listeria* species have two complete sets of DHA PTS (category C). Category D, which includes bacteria that possess a complete DHA PTS plus an extra DhaK, includes only low-G+C gram-positive bacteria. Possibly, the extra DhaK protein allows phosphorylation of an alternative substrate. It is always encoded by a gene distantly related to that encoding DhaK carried within the *dha* operon, which also encodes the remaining proteins of the complete DHA system.

Category E systems, with complete PEP- and ATP-dependent DHA phosphorylating systems, are from two high-G+C gram-positive bacteria. Category F, with just one α-proteobacterial representative, has complete PEP- and ATP-dependent systems but also has two extra DhaKs (GI 13476064 and GI 13488187) and one extra DhaL (GI 13476063). One of the DhaK homologues (GI 13488187) of *M. loti* is encoded on plasmid pMLa.

Several bacteria lack a complete DHA PTS but nevertheless encode within their genomes at least one PTS Dha homologue, as revealed by the data in Table 11. Several different combinations of these proteins can be found. For example, two organisms each have only DhaM (category A) or DhaL (category B). No organism has just DhaK. Just three organisms have only ATP-dependent DhaKL (category C). The DhaKL protein (GI 17937250) of *A. tumefaciens* is encoded on the linear chromosome. Several bacteria (category D) have both DhaK and DhaL, and one of these organisms (*Brucella melitensis*) has a second copy of DhaL (GI 17986682) as well as a $IIA^{Man}$ homologue (GI 17988315) that more closely resembles *E. coli* $IIA^{Man}$ than $IIA^{Dha}$. Category E includes four organisms encoding DhaK, DhaL, and DhaKL, and three of these organisms have a $IIA^{Man}$ homologue. All four genes encoding the Dha proteins in *Sinorhizobium meliloti* were found on the plasmid pSymB. In both *B. mallei* and *B. pseudomallei*, the DhaK and DhaL proteins are encoded on chromosome 2, while the DhaKL protein is encoded on chromosome 1 along with the other PTS homologues found in these organisms. Interestingly, *Bacillus cereus*, which lacks the orphan $IIA^{Man}$ homologue, has a short DhaL fragment (GI 52144352). Finally, organisms in category F, all *Bacillus* species, have DhaKL as well as DhaK. It seems likely that many of these orphan PTS Dha proteins have resulted from genome minimalization, but the possibility that some of them serve a specific function, perhaps to allow phosphorylation of a novel substrate or to provide a regulatory function, should be kept in mind.

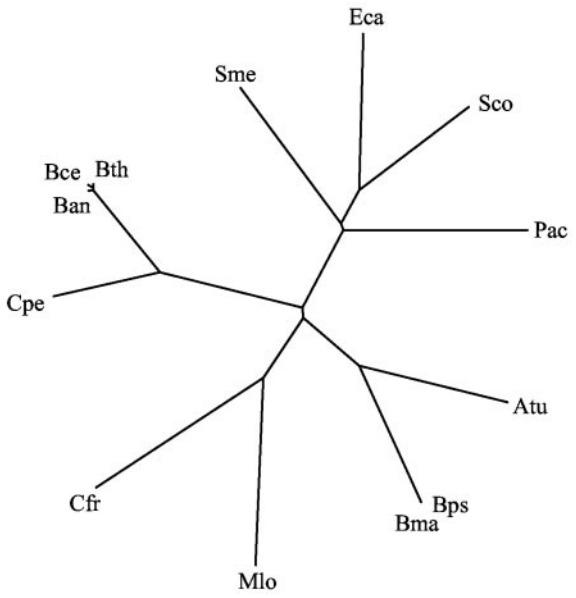



FIG. 4. Phylogenetic tree for the putative ATP-dependent DHA kinases. The DhaKL fusion proteins from 12 organisms (Tables 10 and 11) and the known ATP-dependent dihydroxyacetone kinase from *Citrobacter freundii* (Cfr) were aligned using the Clustal X program, and neighbor-joining trees were generated (102). Trees were viewed using the TreeViewPPC program (113). When multiple strains of a species had been sequenced, only one orthologue was included.

Of the organisms listed in Table 10, one organism, *Bradyrhizobium japonicum*, has an unusual gene arrangement. The complete DHA PTS enzyme II complex as well as enzyme I (GI 27378686) and HPr (GI 27378685) is encoded by adjacent genes that are in an operon, *dhaLMHIXK*. The *B. japonicum dhaM* gene encodes only the $IIA^{Dha}$ domain (GI 27378684). *dhaH* and *dhaI* code for HPr and enzyme I homologues, respectively, while *dhaX* is a hypothetical gene. It is also unusual that *dhaK* and *dhaL* are not adjacent genes. It is interesting that the gene order, *dhaMHI*, is the same as the domain order in the fused tridomain DhaM of *E. coli* and that *B. japonicum* possesses no other PTS enzyme II complexes. In all other organisms listed in category A of Table 10, the *ptsH* and *ptsI* genes map separately from the *dha* operon.

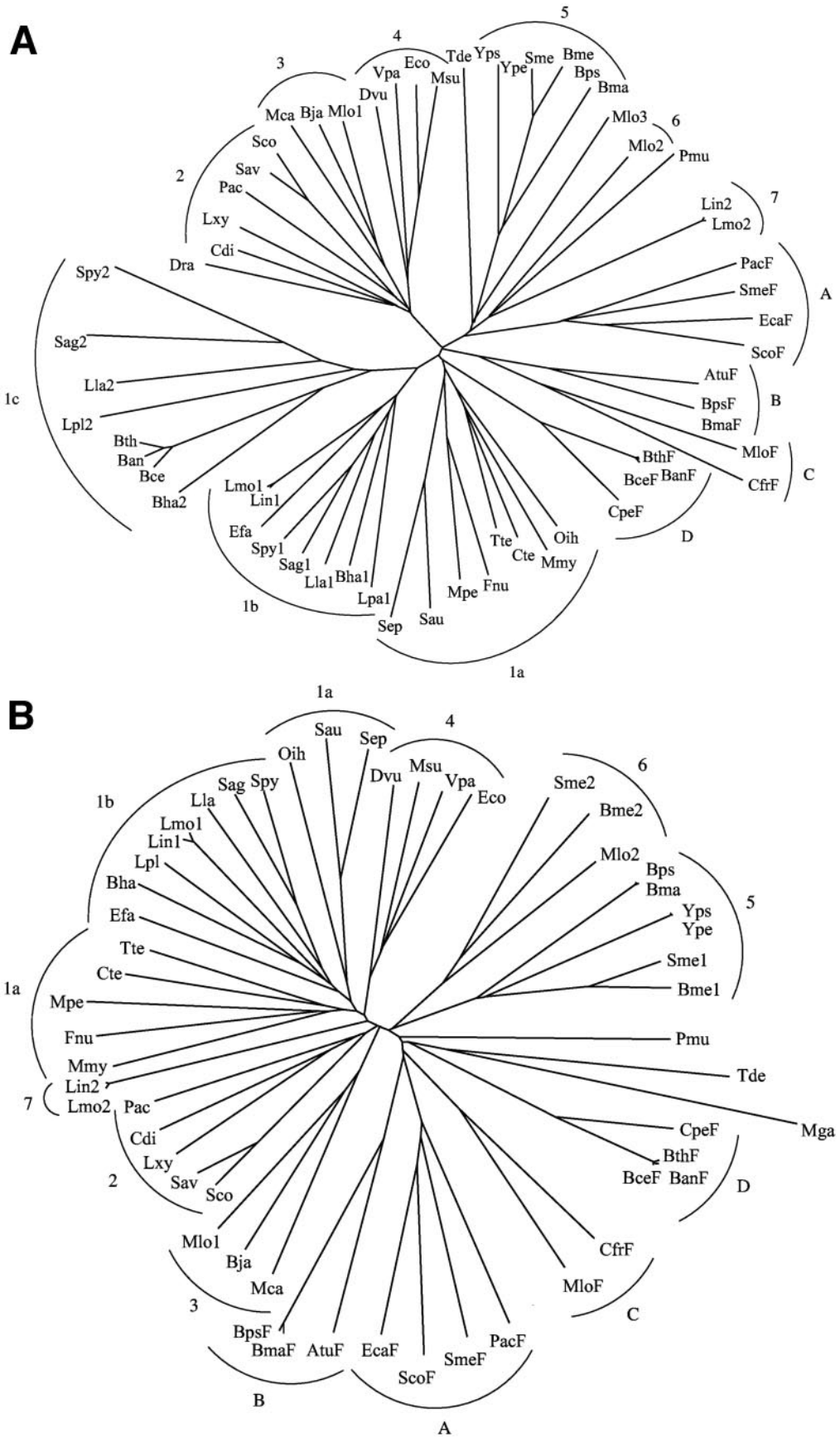
Of the organisms listed in Table 11, *Bartonella quintana* has DhaM (GI 49473737) and several PTS energy-coupling proteins, but no complete enzyme II complex (Table 7); *Brucella melitensis* has DhaK (GI 17986679), DhaL (GI 17986680), and a $IIA^{Man}$-like homologue (GI 17988315), as well as several PTS energy-coupling proteins, but no $IIA^{Dha}$ and no PTS permeases; and *Methylococcus capsulatus* has DhaK (GI 53802782) and DhaL (GI 53802783), but no other PTS protein of any kind. It seems likely that none of these organisms can phosphorylate DHA with either PEP or ATP, but conceivably *M. capsulatus* might use ATP with DhaK and DhaL in a split DhaKL system. We tentatively suggest that some of these organisms are "in transition," losing or gaining complete systems, either by genome reduction or by horizontal transfer, respectively (51, 108).

## PHYLOGENY OF DHA PROTEINS

The Dha proteins were analyzed by generating phylogenetic trees. This was done in several ways. First, a tree was generated just for the fused DhaKL proteins. The results showed that these proteins did not cluster according to organismal group, suggesting either that horizontal transfer of these genes has occurred or that rates of evolutionary sequence divergence have been radically different for the different organisms (11) (Fig. 4). Three separate trees were generated, for the DhaK (Fig. 5A), DhaL (Fig. 5B), and DhaM (Fig. 5C) proteins, for organisms that possess homologues of these proteins (Tables 10 and 11). The analyses clearly suggest that these three proteins have evolved in parallel without shuffling of constituents between systems. However, the 16S rRNA phylogeny results for these organisms (data not shown) suggested that horizontal transfer of complete systems has occurred repeatedly and that different rates of sequence divergence in the different organisms (11) are unlikely to explain the results. The lone DhaM protein (GI 15901038) in *Streptococcus pneumoniae* (Table 11) clusters with the other streptococcal DhaMs (Fig. 5C), but those from *Clostridium perfringens* (GI 18311611) and *Bartonella quintana* (GI 49473737) cluster very loosely together and distantly from all other DhaMs. These two proteins resemble $IIA^{Man}$ more closely than the other DhaM proteins. While *C. perfringens* has a complete mannose enzyme II complex in addition to the orphan $IIA^{Man}$-like protein, *B. quintana* lacks all PTS permeases, including the mannose enzyme II complex.

A few organisms have "extra" DhaLs. These include *S. meliloti* (GI 16264045) (Table 11, category E), *Brucella melitensis* (GI 17986682) (Table 11, category D), and both *M. gallisepticum* (GI 31544231) and *P. syringae* (GI 28870011) (Table 11, category B). Of these organisms, all have DhaK and two DhaLs, except *M. gallisepticum* and *P. syringae*, which have only an orphan DhaL. The extra DhaLs in *S. meliloti* and *B. melitensis* cluster together (Fig. 5B, cluster 6), separate from their paralogues (cluster 5). The *M. gallisepticum* protein is by itself, and it is distantly related to all other homologues (Fig. 5B). Extra DhaKs are present in many low-G+C gram-positive bacteria (Table 10, category D, and Table 11, category F) and one α-proteobacterium, *M. loti* (Table 10, category F). The extra *M. loti* protein (Mlo3) (GI 13488187) is by itself (between clusters 5 and 6) (Fig. 5A), while all extra DhaKs from low-G+C gram-positive bacteria cluster together (cluster 1c). We conclude that these extra DhaL and DhaK proteins did not arise by recent gene duplication events. Instead, they either arose early to fulfill a specific but unknown function or became "orphans" as a result of the genetic loss of their partner proteins.

In categories D and E in Table 11, the DhaL and DhaK proteins are present, but no DhaM homologue could be found. All of the DhaL-DhaK pairs were derived from proteobacteria. Most of these proteins cluster together (cluster 5) in both the DhaL (Fig. 5B) and DhaK (Fig. 5A) trees. They do not cluster with the ATP-dependent DhaKL kinases (clusters A to D). These observations led us to suggest that these DhaL-DhaK protein pairs have coevolved from a common ancestor and serve a unified but unknown function. As noted above, it is possible that they use ATP or another phosphoryl donor to phosphorylate DHA. Alternatively, they could act on another substrate. It is interesting that the Mlo2 DhaL protein (GI 13476063) clusters with Sme2 (GI 16264045) and Bme2 (GI


A
5
4
3
Tde
Yps
Sme
Bme
Bps
Bma
Eco
Vpa
Msu
Dvu
Mca
Bja
Mlo1
Mlo3
6
Mlo2
Pmu
Sco
Sav
2
Pac
Lxy
Cdi
Dra
Lin2
7
Lmo2
Spy2
PacF
SmeF
A
EcaF
ScoF
Sag2
AtuF
Lla2
BpsF
B
1c
Lpl2
BmaF
Bth
Ban
MloF
BthF
Bce
BceF
BanF
CfrF
C
Bha2
Lmo1
Lin1
CpeF
Efa
Spy1
Tte
Oih
D
Sag1
Cte
Mmy
Lla1
Bha1
Mpe
Fnu
Lpa1
1b
Sep
Sau
1a
B
1a
Sau
Sep
4
Oih
Dvu
Msu
Vpa
Eco
6
Sme2
1b
Sag
Spy
Bme2
Lla
Lmo1
Mlo2
Lin1
Bps
Lpl
Bma
5
Bha
Yps
Efa
Ype
Tte
Sme1
Cte
1a
Bme1
Mpe
Pmu
Fnu
Tde
Mmy
Lin2
Mga
7
Lmo2
Pac
CpeF
Cdi
BthF
Lxy
BceF
BanF
2
Sav
D
Sco
Mlo1
CfrF
Bja
Mca
MloF
3
C
BpsF
PacF
BmaF
AtuF
EcaF
SmeF
ScoF
B
A

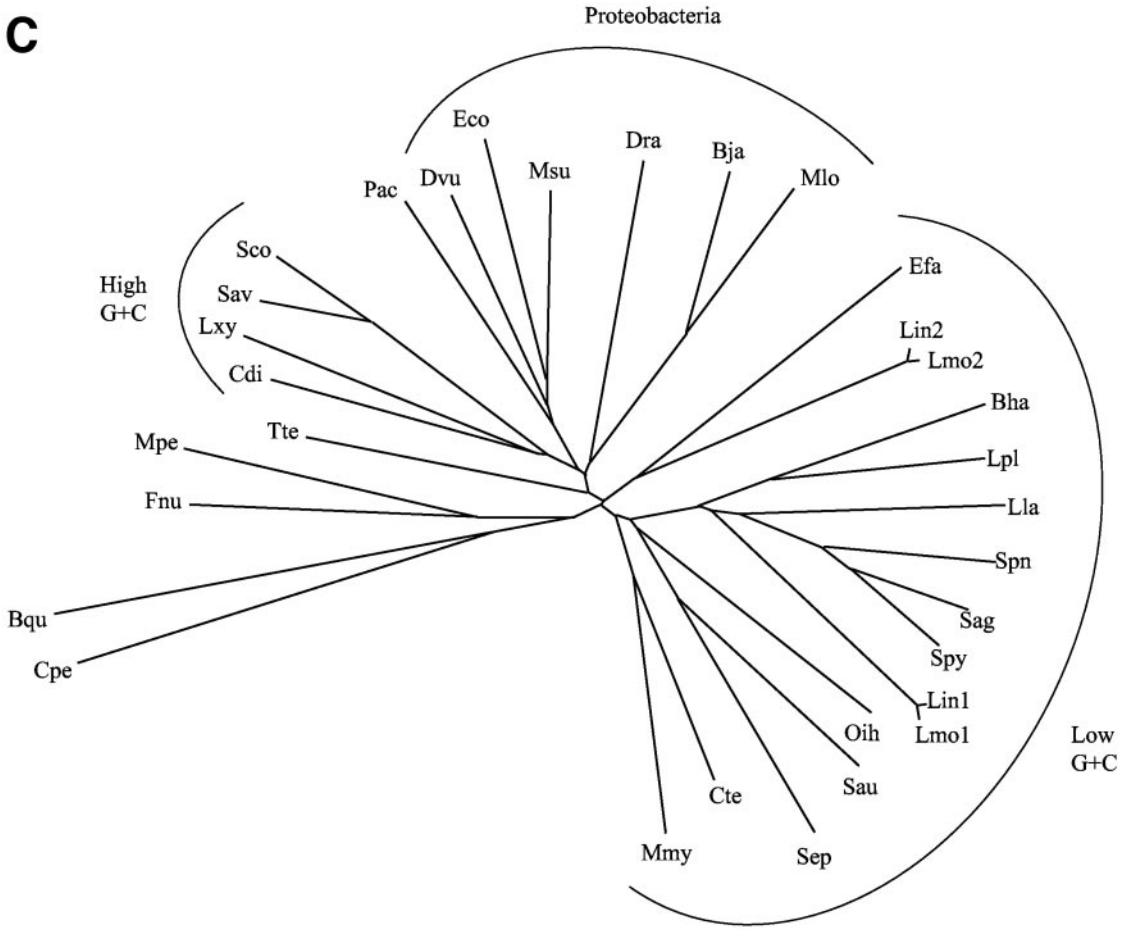



FIG. 5. Phylogenetic trees for DhaK (A), DhaL (B), and DhaM (C) homologues. The trees were generated as described in the legend to Fig. 4. Multiple paralogues from a single organism are distinguished by numbers at the ends of the names. For the DhaK and DhaL trees, the DhaKL fusion proteins were split into the DhaK and DhaL domains and were included in the two trees, respectively. These domains are indicated with an "F" at the ends of the names, and their clusters are designated with letters (A, B, C, and D), while the remaining homologues were grouped numerically into seven clusters. The alphabetical and numerical cluster designations were maintained in the two trees shown in panels A and B. The Clustal X program (102) was used to derive the trees. Abbreviations of organism names are listed in Tables 10 and 11.

17986682) (Fig. 5B). The corresponding Mlo2 DhaK protein (GI 13476064) (Fig. 5A) is by itself because there is no corresponding orthologue in *S. meliloti* and *B. melitensis*.

## PTS DOMAIN FUSION PROTEINS

Table 12 presents the types of PTS domain fusions as well as single-domain proteins that were identified following analysis of the bacterial genomes. These were tabulated according to fusion protein type and family of enzyme II complexes. It is apparent that the largest number of domain combinations was found for the Fru family (13), with the Glc (10) and Man (6) families coming in second and third places, respectively. Only in the Gat family were domain fusions completely lacking (33). In each of the Lac and Gut families, only a single fusion type was found, i.e., IICB in the Lac family and IIBC in the Gut family.

Both bi- and tridomain proteins, including EI and/or HPr, were found. An EI-IIA fusion (Fru family), but no IIA-EI fusion, was found. Surprisingly, no fused protein containing just EI and HPr was identified, although both HPr-IIA and IIA-HPr fusions were found. The former was recognized only in the Fru family, but the latter occurred in both the Fru and Dha families. These observations came as a surprise because tridomain proteins (HPr-EI-IIA [Fru family] and IIA-HPr-EI [Glc, Fru, and Dha families]) were found. In these proteins, HPr is always N-terminal to EI. In no case was HPr C-terminally linked to EI, nor was the IIA domain ever centrally located.

Unfused IIA, IIB, and IIC domains were identified in all PTS families except the Gut family, where an unfused IIB domain could not be found. This is an unusual case because only in the Gut family is the transmembrane IIC protein split into two separate polypeptide chains, i.e., IIC1, with four or five putative transmembrane segments, and IIC2, with three putative transmembrane segments. In all Gut systems, B is fused N-terminally to C1 (encoded by *gutE*), while C2 (encoded by *gutA*) is present as a distinct unfused polypeptide chain. $IIA^{Gut}$ (encoded by *gutB*) is never fused to any other protein domain. Homology between the glucitol IIC proteins and other PTS IIC domains has not been demonstrated (13, 71, 111).

As noted above, only the Man family includes a IID domain. While the Man enzyme II complex usually has all four of its

TABLE 12. PTS fusion proteins that contain only PTS protein domains

| Domain fusion | Occurrence of domain fusion in PTS enzyme II complex family[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Glc | Fru | Lac | Gut[b] | Gat | Man | Asc | Dha[c] |
| Domain fusion combinations that were found | | | | | | | | |
| EI-IIA | | 1 | | | | | | |
| HPr-IIA | | 3 | | | | | | |
| IIA-HPr | | 10 | | | | | | 2 |
| HPr-EI-IIA | | 7 | | | | | | |
| IIA-HPr-EI | 8 | 10 | | | | | | 3[d] |
| IIA | 54 | 101 | 76 | 20 | 25 | 81 | 47 | 26 |
| IIB | 15 | 26 | 81 | | 25 | 56 | 45 | 42 |
| IIC | 7 | 30 | 98 | 16 | 31 | 83 | 49 | 47 |
| IID | | | | | | 83 | | |
| IIAB | | | | | | 28 | 4 | |
| IIBA | 1 | | | | | | | |
| IIBC | 75 | 37 | | 16 | | | | |
| IICB | 80 | 21 | 8 | | | | 7 | |
| IICD[e] | | | | | | 2 | | |
| IIABC | 4 | 47 | | | | | | |
| IIACB | | 3 | | | | | | |
| IIBCA | 56 | 4 | | | | | | |
| IICBA | 36 | 15 | | | | | | |
| Domain fusion combinations that were not found | | | | | | | | |
| EI-HPr | | | | | | | | |
| HPr-EI | | | | | | | | |
| IIA-EI | | | | | | | | |
| EI-HPr-IIA | | | | | | | | |
| EI-IIA-HPr | | | | | | | | |
| HPr-IIA-EI | | | | | | | | |
| IIA-EI-HPr | | | | | | | | |
| IIAC | | | | | | | | |
| IIBAC | | | | | | | | |
| IICA | | | | | | | | |
| IICAB | | | | | | | | |

[a] Glc, Fru, Lac, Gut, Gat, Man, Asc, and Dha indicate the glucose, fructose, lactose, glucitol, galactitol, mannose, ascorbate, and dihydroxyacetone PTS families, respectively. Numbers indicate the numbers of each domain found in the 106 bacterial species (from Tables 7 and 9) that possess PTS homologues. The domain order is as indicated in the first column.

[b] Only in the Gut family is the transmembrane IIC domain split into two polypeptide chains, IIB-IIC1 (GutE; four to five TMSs) and IIC2 (GutA; three TMSs). This unusual arrangement is found in all IIBC homologues except *Chloroflexus auranticus,* where IIBC1 is split into IIB and IIC1. The complete genome of *C. auranticus* was not available at the time of our analyses. One organism, *Lactobacillus johnsonii,* has an orphan $IIA^{Gut}$, while three organisms, *Enterococcus faecalis, Lactobacillus plantarum,* and *Propionibacterium acnes,* have a complete glucitol PTS plus an extra $IIA^{Gut}$. Only *L. plantarum* and *Salmonella enterica* serovar Typhimurium have two complete enzyme II complexes in the glucitol family.

[c] In the Dha family, the DhaM, DhaL, and DhaK proteins correspond to the $IIA^{Dha}$, $IIB^{Dha}$, and $IIC^{Dha}$ proteins, respectively.

[d] Only one organism, *Desulfovibrio vulgaris,* possesses a full-length $IIA^{Dha}$-HPr-EI fusion, while in two organisms, *E. coli* and *S. flexneri,* the C-terminal region of EI is truncated.

[e] The IID domain in the Man family was found fused to IIC in the order IICD but was not present in any other fusion combination.

protein constituents unfused, IIA can be fused N-terminally to IIB, and IIC can be fused N-terminally to IID. No other fusion combinations in the Man family were detected. Thus, although A-B fusions occur, B-A fusions do not, and although C-D fusions occur, D-C fusions could not be found. Moreover, A and B are never fused to C or D (Table 12).

While A-B fusions were found only in the Man and Asc families, B-A fusions were found only in the Glc family. In contrast, both B-C and C-B fusions were identified in both the Glc and Fru families, and the latter fusion type was also found in the Lac and Asc families.

Tridomain PTS permease proteins with IIA, IIB, and IIC fused in single polypeptide chains were restricted to the Glc and Fru families, and four of the six possibilities were found. Thus, ABC, BCA, and CBA were identified in both families, and ACB was found in the Fru family, but BAC and CAB were not found. It is surprising that A can be fused either N-terminally or C-terminally to C in a tridomain protein. Interestingly, this only occurs when B is directly fused to C, but not to A. These two domains (A and C) were never found linked in a bidomain protein. While stereospecific constraints may explain some of these observations, others seem inexplicable (see "Conclusions").

Some of the fusions were found only in certain taxonomic groups. For example, the EI-IIA, HPr-IIA, IIA-HPr, HPr-EI-IIA, and IIA-HPr-EI proteins in the Fru family as well as IIA-HPr-EI fusions in the Glc and Dha families were found only in *Proteobacteria*. The two IIA-HPr fusions found in the Dha family were found in two *Actinobacteria*, *Corynebacterium diphtheriae* (GI 38234870) and *Leifsonia xyli* (GI 50954678). A single EI-$IIA^{Fru}$ protein (GI 15804544) was found in only one of the four strains of *E. coli* analyzed, strain O157:H7 EDL933. This fusion could have resulted from a sequencing error that split the HPr-EI-$IIA^{Fru}$ protein found in the other three strains of *E. coli*. The HPr-EI-$IIA^{Fru}$ fusion was found in several γ-*Proteobacteria* and one member of the β-*Proteobacteria*, *Chromobacterium violaceum* (GI 34497766). Three HPr-$IIA^{Ntr}$ (Fru family) fusions were found in the three *Vibrio* species analyzed. Although $IIA^{Fru}$-HPr fusions were found only in γ-*Proteobacteria*, $IIA^{Fru}$-HPr-EI fusions were found in α-, β-, and γ-*Proteobacteria*. $IIA^{Glc}$-HPr-EI fusions were found mostly in β-*Proteobacteria*, but also in *Caulobacter crescentus* (α-*Proteobacteria*; GI 16124703 and 16124792) and *P. aeruginosa* (γ-*Proteobacteria*; GI 15598955).

The $IIBA^{Glc}$ fusion was found only in *Staphylococcus aureus* (GI 49482502). Since this fusion was found in all five strains of *S. aureus* analyzed, it is unlikely to be a sequencing artifact. $IIABC^{Glc}$ fusions were found (as two copies) only in the two *Lactobacillus* species analyzed. $IIAB^{Asc}$ fusions were found only in the *Actinobacteria*. Two $IICD^{Man}$ fusions were found in two thermophilic bacteria, *Symbiobacterium thermophilum* (GI 51892416) and *Thermoanaerobacter tengcongensis* (GI 20806720). The $IICB^{Lac}$ fusions were found only in a few *Firmicutes*, specifically in *Staphylococcus* species, *Streptococcus* species, and *Clostridium acetobutylicum* (GI 15895217). While the $IIBCA^{Fru}$ fusions were found only in the *Firmicutes*, the $IIACB^{Fru}$ fusion was found only in the three *Vibrio* species (γ-*Proteobacteria*). All other fusions were found distributed in diverse taxonomic groups, although interestingly, the $IIBC^{Fru}$ and $IICBA^{Fru}$ fusions predominated in *Proteobacteria* while the $IIBCA^{Glc}$ and $IIABC^{Fru}$ fusions were found mostly in *Firmicutes* and *Actinobacteria*. $IICBA^{Glc}$ fusions also predominated in the *Firmicutes*.

## PTS PROTEINS WITH EXTRA NON-PTS DOMAINS

Transcriptional activators and antiterminators of the PTS regulatory domain (PRD) family are known to incorporate

TABLE 13. PTS proteins with extra non-PTS domains[a]

| Category | Protein | Protein size (aa) | Length of X (aa) | Homology of X | Organism | GI no. |
|---|---|---|---|---|---|---|
| A | $IIA^{Fru}$-X-FPr-FPr | 499 | 150 | Unknown | *Haemophilus influenzae* Rd KW20 | 16272396 |
| | $IIA^{Fru}$-X-FPr-FPr | 502 | 150 | Unknown | *Mannheimia succiniciproducens* MBEL55E | 52426235 |
| B | $IIA^{Fru}$-X | 271 | 120 | Unknown | *Chlorobium tepidum* TLS | 21673462 |
| | $IIA^{Ntr}$-X | 311 | 150 | CBS domain | *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586 | 19705231 |
| | $IIABC^{Glc}$-X | 841 | 170 | MAD, Smc | *Mesoplasma florum* L1 | 50364823 |
| | $IIABC^{Glc}$-X | 858 | 180 | Sbc | *Mesoplasma florum* L1 | 50364848 |
| | $IIABC^{Glc}$-X | 874 | 170 | HEC1 | *Mesoplasma florum* L1 | 50365435 |
| | $IIABC^{Glc}$-X | 849 | 200 | Zn ribbon | *Mesoplasma florum* L1 | 50365128 |
| | $IIABC^{Glc}$-X | 854 | 200 | Smc | *Mesoplasma florum* L1 | 50365134 |
| | $IIABC^{Glc}$-X | 1,020 | 280 | Smc | *Mycoplasma penetrans* HF-2 | 26553906 |
| | $IIBC^{Glc}$-X | 628 | 130 | TOPEUc, Smc | *Mycoplasma mycoides* subsp. *mycoides* SC strain PG1 | 42561174 |
| | $IIBC^{Fru}$-X | 613 | 80 | Unknown | *Mycoplasma mycoides* subsp. *mycoides* SC strain PG1 | 42561366 |
| | $IIC^{Asc}$-X | 660 | 135 | Unknown | *Mycoplasma pneumoniae* M129 | 13508235 |
| C | X-$IIA^{Ntr}$ | 225 | 80 | HTH motif | *Chlamydia trachomatis* D/UW-3/CX | 15605011 |
| | X-$IIA^{Ntr}$ | 226 | 75 | HTH motif | *Chlamydophila caviae* GPIC | 29840113 |
| | X-$IIA^{Ntr}$ | 225 | 75 | HTH motif | *Chlamydophila pneumoniae* AR39 | 16752983 |
| | X-$IIA^{Ntr}$ | 239 | 90 | HTH motif | *Parachlamydia* sp. strain UWE25 | 46445900 |
| | X-$IIA^{Ntr}$ | 235 | 90 | HTH motif | *Pirellula* sp. strain 1 | 32473906 |
| | X-$IIA^{Ntr}$ | 225 | 75 | Unknown | *Geobacter sulfurreducens* PCA | 39995841 |
| | X-$IIA^{Ntr}$ | 201 | 60 | Unknown | *Treponema denticola* ATCC 35405 | 42526589 |
| | X-$IIA^{Gat}$ | 215 | 55 | Unknown | *Lactobacillus johnsonii* NCC 533 | 42518211 |
| | X-$IIB^{Fru}$ | 414 | 320 | Unknown | *Mycoplasma gallisepticum* R | 31544567 |
| | X-$IIBC^{Fru}$ | 564 | 110 | Unknown | *Mannheimia succiniciproducens* MBEL55E | 52426233 |
| D | X-$IIC^{Glc}$ | 492 | 165 | Unknown | *Salmonella enterica* subsp. *enterica* serovar Typhi Ty2 | 29144387 |
| | X-$IIC2^{Gut}$ | 230 | 50 | Unknown | *Mesorhizobium loti* MAFF303099 | 13473118 |
| E | X-$IIC^{Fru}$-X | 629 | N-100, C-150 | Unknown, unknown | *Mycoplasma mycoides* subsp. *mycoides* SC strain PG1 | 42560835 |
| | X-$\Delta IIB^{Asc}$ | 178 | 80 | Unknown | *Salmonella enterica* subsp. *enterica* serovar Typhi Ty2 | 29144680 |
| | $\Delta IIB^{Glc}$-X | 87 | 40 | Unknown | *Lactobacillus johnsonii* NCC 533 | 42519701 |
| F | $IIC^{Lac}$-X | 704 | 250 | DUF2 | *Vibrio cholerae* O1 bv. *eltor* strain N16961 | 15641224 |
| | $IIC^{Lac}$-X | 698 | 250 | DUF2 | *Vibrio vulnificus* CMCP6 | 27366325 |
| | $IIC^{Lac}$-X | 655 | 250 | DUF2 | *Vibrio vulnificus* CMCP6 | 27367219 |
| | $IIC^{Lac}$-X | 681 | 250 | DUF2 | *Vibrio parahaemolyticus* RIMD 2210633 | 28901284 |
| | $IIC^{Lac}$-X | 704 | 250 | DUF2 | *Wolinella succinogenes* DSM 1740 | 34557244 |
| G | X-$IIA^{Fru}$ | 589 | 415 | $Na^+/H^+$ antiporter | *Pirellula* sp. strain 1 | 32474980 |
| | X-$IIA^{Fru}$ | 701 | 550 | $Na^+/H^+$ antiporter | *Borrelia garinii* PBi | 51598702 |
| | X-$IIA^{Fru}$ | 701 | 550 | $Na^+/H^+$ antiporter | *Borrelia burgdorferi* B31 | 15594792 |
| | X-$IIA^{Glc}$ | 648 | 500 | $Na^+$/melibiose symporter | *Lactobacillus plantarum* WCFS1 | 28379785 |
| | X-$IIA^{Glc}$ | 652 | 500 | $Na^+$/melibiose symporter | *Lactobacillus plantarum* WCFS1 | 28379802 |
| | X-$IIA^{Glc}$ | 653 | 500 | $Na^+$/melibiose symporter | *Lactobacillus johnsonii* NCC 533 | 42518347 |
| | X-$IIA^{Glc}$ | 654 | 500 | $Na^+$/melibiose symporter | *Lactobacillus johnsonii* NCC 533 | 42518782 |
| | X-$IIB^{Glc}$ | 349 | 265 | TIM | *Pasteurella multocida* Pm70 | 15603505 |
| | X-X-$IIA^{Fru}$ | 967 | N-475, 350 | PspF, COG3933 | *Thermoanaerobacter tengcongensis* | 20806706 |

[a] Proteins were grouped for convenience based on (i) whether the non-PTS domains were partial domains (categories A to E) or full-length domains (categories F and G) and (ii) whether the non-PTS domain was in the middle (category A), at the C terminus (categories B and F), or at the N terminus (categories C to E) of the protein. While category C includes proteins with non-PTS domains fused N-terminally to PTS enzyme IIA or IIB, category D includes N-terminal fusions of non-PTS domains to the PTS enzyme IIC; category E is a miscellaneous group of fusion proteins.

PTS IIA and IIB domains into their structures (24, 29). Excluding these known cases where PTS and non-PTS domains are fused, we have tabulated all other examples revealed by our genome analyses. Domains were identified by searching the conserved domain database (42). The types of fusions identified are listed in Table 13. The proteins are grouped according to the positions of the extra domains (indicated with X's in Table 13).

As reported previously, the FruB proteins of *E. coli* and related organisms have a IIA-X-FPr domain structure (67). In *Haemophilus* and *Mannheimia*, the same domain order is found, but an extra C-terminal FPr (fructose-inducible HPr) is present (69), probably due to a recent intragenic duplication event (category A in Table 13). In *Chlorobium* and *Fusobacterium* (category B), $IIA^{Fru}$-X and $IIA^{Ntr}$-X proteins of similar sizes are found. These are homologous to each other throughout their lengths, and the IIA domains are homologous to the IIA domain of the FruB protein of *E. coli*. In *F. nucleatum*, the extra domain is homologous to CBS domains found in membrane proteins that are probably involved in signal transduction (COG34481).

Other PTS proteins with C-terminal X domains are from organisms within the *Mycoplasma* group (category B). The C-terminal X domains in these proteins share homology with proteins that interact with DNA, such as the MAD (*m*itotic *a*rrest *d*eficient or mitotic checkpoint) proteins (107), the Smc (*s*tructural *m*aintenance of *c*hromosomes) proteins (34), the SbcC protein (an ATPase involved in DNA repair) (15), the HEC1 (*h*ighly *ex*pressed in *c*ancer) protein, which may play a role in chromosomal segregation (43), and the TOPEUc (DNA *top*oisomerase 1 [*Eukaryota*]) protein (92). All of these homologous domains are linked to $IIC^{Glc}$ domains in various PTS permeases (Table 13). It is possible that some of these X domains function in DNA binding or in protein-protein interactions. The PTS proteins to which these X domains in category B are fused include $IIABC^{Glc}$ proteins as well as $IIBC^{Glc}$, $IIBC^{Fru}$, and $IIC^{Asc}$ proteins.

X domains are fused N-terminally to $IIA^{Ntr}$ domains in bacteria of the chlamydial kingdom, in the δ-proteobacterium *Geobacter sulfurreducens*, and in the spirochete *Treponema denticola* (category C in Table 13). At least some of these domains are homologous to helix-turn-helix DNA binding domains in transcriptional regulators of the MerR family (8). It seems likely that these proteins function in DNA binding, possibly to regulate gene expression in response to PTS $IIA^{Ntr}$ domain phosphorylation. In other fusions, the X domains, of variable size and sequence, are fused to $IIA^{Gat}$, $IIB^{Fru}$, and $IIBC^{Fru}$ proteins.

In category D, we found X domains linked N-terminally to the integral membrane $IIC^{Glc}$ or $IIC^{Gut}$ domain. In the *Salmonella* $IIC^{Glc}$ protein, the X domain is 165 residues long, but in the $IIC2^{Gut}$ homologue from *M. loti*, it is only 50 residues long.

In category E (miscellaneous), we found X domains linked both N- and C-terminally to a partially homologous $IIC^{Fru}$ protein. This could be a sequence-divergent IIBCA (Table 12). Finally, there is an $X\Delta IIB^{Asc}$ (X = 80 residues) and a $\Delta IIB^{Glc}$-X (X = 40 residues) protein where the IIB domains are N-terminally truncated. Because of the truncations, it is unlikely that these proteins function as classical PTS phosphoryl transfer proteins.

*Vibrio* species and *Wolinella succinogenes* have $IIC^{Lac}$ homologues fused to C-terminal DUF2 domains (category F). The DUF2 domain, now characterized as the EAL domain, occurs ubiquitously in bacteria and has been shown to be a characteristic feature of cyclic diguanylate-specific phosphodiesterases (89, 95). Moreover, putative $Na^+/H^+$ antiporters of the monovalent cation-proton antiporter 2 family (TC no. 2.A.37) are linked to $IIA^{Fru}$ domains in *Pirellula* and *Borrelia* species (category G). C-terminal $IIA^{Glc}$-like domains in β-galactoside permeases of the GPH family (TC no. 2.A.2) are found in lactobacilli, where they serve a regulatory function (60, 61). In *Pasteurella multocida*, a triosephosphate isomerase, a glycolytic enzyme involved in the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate, contains a C-terminally fused $IIB^{Glc}$ domain. Finally, in *Thermoanaerobacter tengcongensis*, a transcriptional regulator with an ATPase domain is fused to a C-terminal $IIA^{Fru}$ domain. The domain organization in this protein most closely resembles the domain structure of the LevR protein of *Bacillus subtilis*, except that the $IIB^{Gat}$ domain cannot be recognized, possibly due to extensive sequence divergence, and the C-terminal PRD domain is replaced with a $IIA^{Fru}$-like domain. This structure is reminiscent of an HPr-transcription factor fusion protein identified previously in *Clostridium acetobutylicum*, where an HPr-like domain is fused N-terminally to a similar (in sequence) transcription factor (73).

The results summarized in this section and in Table 13 lead to the suggestion that PTS proteins and their phosphorylation regulate macromolecular interactions such as protein-protein and protein-DNA binding. The identification of these fusion proteins leads to predictions with respect to several novel regulatory functions of the PTS, including the regulation of ion transport, transcription, enzymatic activities, and several types of macromolecular interactions. It is likely that these postulated regulatory functions will prove to augment the known list of PTS functions presented in Table 2. It should be noted that in some cases where non-PTS domains of unknown function are apparently fused to PTS domains, errors in sequencing and ORF prediction could account for artifactual fusions.

## PTS PROTEIN STRAIN DIFFERENCES IN SINGLE BACTERIAL SPECIES

In several cases, the genomes of multiple strains for a single species have been fully sequenced (Table 14). For example, four different strains each of *Bacillus anthracis*, *Chlamydophila pneumoniae*, and *Escherichia coli* and five different strains each of *Staphylococcus aureus* and *Streptococcus pyogenes* have been sequenced. The availability of these genome sequences provides us with the unique capability of comparing the complements of PTS proteins in closely related strains. Seven of the 18 species listed in Table 14 had no differences in PTS protein composition between the different strains for which complete genome data were available. In addition, two species, *B. anthracis* and *S. enterica* serovar Typhi, each exhibits only a single difference, a split gene which could be due to a sequencing error (Table 14). The nine remaining species listed in Table 14 are believed to exhibit true strain differences.

Of the four *E. coli* strains, the three pathogenic strains all possess extra Asc, Gat, and Man systems, as well as an extra HPr homologue, that are lacking in *E. coli* K-12. Additionally, one strain (CFT073) has an extra Lac-type system. The genome sizes of these four *E. coli* strains differ substantially (108). The two strains of *Shigella flexneri*, which are as close to *E. coli* as the different *E. coli* strains are to each other, differ from one another only in possessing a large genomic inversion that carries genes encoding complete Fru, Glc, and Man systems. Consequently, there is no PTS compositional difference between these two strains. They most closely resemble the *E. coli* K-12 strain but differ from this strain in that the *S. flexneri* strains possess an extra Man system. The two strains of *Vibrio vulnificus* differ with respect to their complete Fru and Man systems, and they exhibit inversions and insertions with respect to the arrangement of some of their *pts* genes relative to each other. Only one of the three *Yersinia pestis* strains possesses a split Asc system, with the genes encoding the IIA and IIB proteins together but separated from the IIC protein-encoding gene. In the other two *Y. pestis* strains, these genes are found together in a single putative operon.

Two *Bacillus* species, *B. cereus* and *B. licheniformis*, include strains that differ with respect to the presence or absence of a single PTS permease system, an ascorbate (Asc)-type system for *B. cereus* and a fructose (Fru)-type system for *B. licheniformis*. Of the two strains of *Listeria monocytogenes* that have

TABLE 14. Differences in PTS protein composition between strains of single bacterial species

| Organism and strain | Strain difference(s) |
|---|---|
| Proteobacteria | |
| *Buchnera aphidicola* (*Baizongia pistaciae*) | None |
| *Buchnera aphidicola* APS (*Acyrthosiphon pisum*) | |
| *Buchnera aphidicola* Sg (*Schizaphis graminum*) | |
| *Escherichia coli* K-12 | |
| *Escherichia coli* O157:H7 | Extra Asc, Gat, and Man systems and HPr |
| *Escherichia coli* O157:H7 EDL933 | Extra Asc, Gat, and Man systems and HPr |
| *Escherichia coli* CFT073 | Extra Asc, Gat, Man, and Lac systems and HPr |
| *Shigella flexneri* 2a strain 2457T | Genome rearrangement, a segment of DNA (carrying Fru, Glc, and Man genes) flipped with respect to the other strain |
| *Shigella flexneri* 2a strain 301 | |
| *Salmonella enterica* subsp. *enterica* serovar Typhi CT18 | Lacks an intact EI (sequencing error?) |
| *Salmonella enterica* subsp. *enterica* serovar Typhi Ty2 | |
| *Vibrio vulnificus* CMCP6 | |
| *Vibrio vulnificus* YJ016 | Several rearrangements and insertions; has extra Fru (IIA and IIBC) and Man (IIA, IIB, IIC, and IID) systems |
| *Yersinia pestis* CO92 | IIB Asc found by DNA BLAST; The Asc operon is split: IIA and IIB are together, but IIC is separate |
| *Yersinia pestis* bv. Medievalis strain 91001 | |
| *Yersinia pestis* KIM | |
| *Xylella fastidiosa* 9a5c | None |
| *Xylella fastidiosa* Temeculal | |
| *Neisseria meningitidis* MC58 | None |
| *Neisseria meningitidis* Z2491 | |
| Gram-positive bacteria | |
| *Bacillus anthracis* A2012 | $IICBA^{Glc}$ is split into a truncated IICBA $(\Delta)^{Glc}$ and $(\Delta)$ $IIA^{Glc}$ (sequencing error?) |
| *Bacillus anthracis* Ames | $IICBA^{Glc}$ |
| *Bacillus anthracis* Ames ancestor | $IICBA^{Glc}$ |
| *Bacillus anthracis* Sterne | $IICBA^{Glc}$ |
| *Bacillus cereus* 10987 | |
| *Bacillus cereus* 14579 | Lacks an Asc system (IIA, IIB, and IIC) |
| *Bacillus cereus* ZK | |
| *Bacillus licheniformis* ATCC 14580 | |
| *Bacillus licheniformis* DSM 13 | Has an extra Fru system (IIA and IICB) |
| *Chlamydophila pneumoniae* AR39 | None |
| *Chlamydophila pneumoniae* CWL029 | |
| *Chlamydophila pneumoniae* J138 | |
| *Chlamydophila pneumoniae* TW-183 | |
| *Listeria monocytogenes* EGD-e | Has two extra Glc (IIBCA), three extra Fru (IIA, IIB, and IIC; IIA and IIBC), and one extra Asc (IIA, IIB, and IIC) system |
| *Listeria monocytogenes* 4b F2365 | |
| *Staphylococcus aureus* subsp. *aureus* MRSA252 | None |
| *Staphylococcus aureus* subsp. *aureus* MSSA476 | |
| *Staphylococcus aureus* subsp. *aureus* Mu50 | |
| *Staphylococcus aureus* subsp. *aureus* MW2 | |
| *Staphylococcus aureus* subsp. *aureus* N315 | |
| *Streptococcus agalactiae* 2603V/R | |
| *Streptococcus agalactiae* NEM316 | Has an extra Lac system (IIA and IICB) |
| *Streptococcus pneumoniae* R6 | BglG-$IIA^{Fru}$ is encoded by a single ORF |
| *Streptococcus pneumoniae* TIGR4 | Has an extra Fru (IIA, IIB, and IIC) system as well as $IIA^{Ntr}$; BglG and $IIA^{Fru}$ are encoded by separate ORFs |
| *Streptococcus pyogenes* M1 GAS | |
| *Streptococcus pyogenes* MGAS315 | |
| *Streptococcus pyogenes* SSI-1 | |
| *Streptococcus pyogenes* MGAS8232 | Has DhaKLM genes and an additional DhaK protein |
| *Streptococcus pyogenes* MGAS10394 | Has DhaKLM genes, and the additional DhaK protein is split |
| Spirochetes | |
| *Leptospira interrogans* serovar Copenhageni strain Fiocruz L1–130 | None |
| *Leptospira interrogans* serovar Lai strain 56601 | |

had their genomes sequenced, one (EGD-e) has two extra Glc systems, three extra Fru systems, and one extra Asc system compared to the other (4bF2365).

The three species of *Streptococcus* listed in Table 14 also show strain differences. The two *Streptococcus agalactiae* strains differ with respect to the presence or absence of a complete Lac system. The five *S. pyogenes* strains differ from each other in that two have an extra complete Dha system as well as an extra orphan DhaK protein that is lacking in the other three strains. The two sequenced *S. pneumoniae* strains differ in that one possesses a complete Fru system and a $IIA^{Ntr}$ protein that are lacking in the other strain. Additionally, the R6 strain of *S. pneumoniae* has a fused BglG-$IIA^{Fru}$ protein (GI 15902323) which is split into two separate proteins, BglG (GI 15900239) and $IIA^{Fru}$ (GI

15900240), in strain TIGR4. Whether this difference is due to a sequencing error cannot be determined with certainty using bioinformatic approaches alone.

In each case where strain differences were observed for a single species, particularly where genes encoding PTS proteins were either present or lacking, one can conclude that *pts* genes have been gained and lost relatively frequently during recent evolutionary history. This observation suggests that these genes have been transferred horizontally with a high frequency. This conclusion is substantiated by phylogenetic relationships which argue against a strict vertical descent for several PTS permeases (data not shown). It is interesting that many of the strain differences we observed involve the gain or loss of complete enzyme II complexes, suggesting that the acquisition or loss of these genes has functional significance and was therefore subject to selection.

## CONCLUSIONS

In this study, we examined 202 genomes for homologues of all known PTS proteins. Homologues were found only in bacteria, in agreement with our suggestion that the PTS evolved late, after the three domains of life separated (81). Within the bacterial domain, 22% of the species examined had no PTS protein homologues, 21% had only soluble putative regulatory proteins, and 57% had all constituents required for sugar phosphorylation, including at least one complete PTS permease. Organisms lacking PTS homologues included several of the most primitive bacteria with sequenced genomes (*Aquafex*, *Thermus*, and *Thermotoga*). *Thermotoga maritima* is a saccharolytic bacterium that metabolizes several simple and complex carbohydrates, including glucose, sucrose, maltose, starch, cellulose, and xylose (31, 32). It possesses nine putative sugar-specific ABC transport systems for the uptake of these carbon sources (52). This fact again argues in favor of the late evolution of the PTS. However, all cyanobacteria examined, representative α-, δ-, ε-, and γ-proteobacteria, several actinobacteria, and one mollicute also lacked PTS protein homologues (Table 6). Except for the cyanobacteria, many of these *pts* gene-lacking genomes probably resulted from genome minimalization. Twenty-nine species of bacteria were found to possess only soluble PTS protein homologues, presumed to function in regulation (7, 23). These organisms included the chlamydiae, several spirochetes, a green bacterium, *Chlorobium tepidum*, and several proteobacteria, particularly in the alpha and beta subcategories (Table 7). Nevertheless, bacteria with complete complements of PTS energy-coupling proteins plus permeases could be found in almost all of these bacterial kingdoms, except the chlamydial kingdom and the primitive bacterial kingdoms (Table 9). It would therefore appear on the basis of these observations that PTS protein-encoding genes have been gained and lost with a high frequency.

This last conclusion was substantiated by analyses of various strains of a single species and of various species in a single genus where variations in the complement of PTS permeases varied drastically (Table 14). Thus, for example, different *E. coli* strains possess between 17 and 26 PTS permeases, and the different strains differ with respect to the presence of members of 4 PTS permease families (Asc, Gat, Man, and Lac; Table 14). Similarly, major differences were observed between different strains of *Vibrio vulnificus*, *Listeria monocytogenes*, and several *Streptococcus* species (Table 14). In contrast, several sequenced strains of other species (including *Chlamydophila pneumoniae*, *Staphylococcus aureus*, *Buchnera aphidicola*, *Xylella fastidiosa*, and *Neisseria meningitidis*) exhibited no differences in PTS protein content. We concluded that the gain and loss of PTS permeases has occurred repeatedly, but in a species-specific fashion. This has apparently resulted from genome minimalism (51) as well as the horizontal transfer of genetic information encoding PTS permeases and energy-coupling proteins (108). The detection of genes encoding PTS protein homologues on mobile genetic elements (14, 88, 103, 112; see above) substantiates this last conclusion. Recently, the evolution of the mannose PTS transporters has been discussed, and extensive horizontal transfer of the genetic material encoding these systems has been documented (114).

Within a single coherent genus, different species similarly show differences in PTS permease content. Thus, corynebacterial species may either possess or lack a fructose-type PTS and an ascorbate-type PTS permease. More surprising, *Mycoplasma* species with drastically reduced genome sizes may possess between two and nine complete PTS permeases. Differences were also noted among the *Streptomyces* and *Clostridium* species (Table 9).

In this study, we divided the PTS permeases into seven families (Glc, Fru, Lac, Gut, Gat, Man, and Asc) (Table 9). The occurrence of the members of these families is summarized in Fig. 2. Of the 77 bacterial species analyzed that encode PTS permeases within their genomes, the glucose (Glc) family was most highly represented. The order of prevalence of the seven families was as follows: Glc (30%) > Fru (25%) > Man (15%) > Lac (14%) > Asc (9%) > Gat (4%) > Gut (3%). However, the different taxonomic groups show various proportions of each of the PTS permease families. While the most abundant family in the *Firmicutes* was the Glc family, followed by Fru, Lac, and Man, the most prevalent family in the *Proteobacteria* was the Fru family, followed by Glc, Man, and Asc. When it is considered that the Glc, Fru, and Lac families actually belong to a single superfamily, while the Gat and Asc families combine to form a second superfamily (13, 33, 81), it can be concluded that the Glc/Fru/Lac superfamily includes 69% of all PTS permeases. The Man family includes 15%, the Asc/Gat superfamily includes 13%, and the Gut family includes 3%. This observation is in agreement with our suggestion that the fructose PTS was the first primordial system to have evolved (75, 81). These arguments are strengthened by the finding reported here that more bacteria with a single type of PTS permease have a Fru-type system than any other type.

The physiological functions of HprK homologues have been identified only for the low-G+C gram-positive *Firmicutes*, although many such homologues have been identified in gram-negative bacterial kingdoms (29, 97). We have postulated that all such enzymes serve regulatory functions, but only in a few instances are clues available as to what those functions may be. In the case of α-*Proteobacteria*, truncated HprKs are found in operons with genes encoding the gluconeogenic enzyme PEP carboxykinase, a sensor kinase/response regulator pair, and other PTS proteins such as HPr and IIA homologues (29). We have proposed that the PTS proteins function in a phosphoryl transfer cascade (65) that regulates the expression of the *pck* gene encoding PEP carboxykinase, possibly via the sensor ki-

nase/response regulator pair (7, 29). No experimental data have bearing on this point. The multiple distantly related HprKs found in other α-*Proteobacteria*, such as *Rhodospirillum rubrum*, are also of unknown function (97).

We have concluded that the DHA PTS enzyme II complex is a recently evolved system derived from an ATP-dependent DHA kinase (81). The fact that the DhaM components vary dramatically in their domain compositions argues in favor of this conclusion. Nevertheless, these systems occur in a wide range of bacterial kingdoms (Table 10). This may have resulted from horizontal transfer of the genes encoding these systems, as suggested by our phylogenetic analyses (see below).

Phylogenetic analyses revealed a lack of orthology between PTS Dha proteins from a variety of bacteria. Thus, although an excellent phylogenetic correlation was observed between the DhaK, DhaL, and DhaM trees (Fig. 5), a very poor correlation was observed between these trees and the 16S rRNA trees for the corresponding organisms (data not shown). The implication is that while little or no shuffling of the three constituents of the Dha systems has occurred throughout their evolutionary divergence, they have been transferred laterally together as a unit. While orphan PTS proteins are frequently encoded within bacterial genomes, they may not always be functional. They may be the result of residual inactive genetic information resulting from genome minimalism (51).

The instability of the DHA PTS is further indicated by the variation in the structures of the DhaM components. Some DhaM proteins contain only the $IIA^{Dha}$ domain, but three homologous fusion proteins were also identified. These included (i) a $IIA^{Dha}$-HPr fusion, (ii) a $IIA^{Dha}$-HPr-EI fusion, and (iii) a $IIA^{Dha}$-HPr-EIΔ fusion with the C-terminal region of enzyme I missing. In *E. coli*, which possesses a DhaM protein with a type 3 fusion structure, the classical enzyme I and HPr are required for DhaM phosphorylation. These two energy-coupling proteins may not be required for phosphorylation of the type 2 fusion proteins. Furthermore, in *Bradyrhizobium japonicum*, the *dha* operon encodes DhaM, HPr, and enzyme I with the gene order *dhaMHI*, the same order as that observed for the type 2 and type 3 fusion proteins mentioned above. The type 2 fusion proteins could have resulted from the elimination of chain termination codons and/or the introduction of intragenic microdeletions. It thus seems that the scenario found in *B. japonicum* could represent a transitional state towards production of the tridomain fusion proteins. We may be visualizing a "snapshot" of the evolutionary process still in progress (81).

We have identified many types of previously unidentified PTS protein fusions present in the bacterial genomes analyzed. EI-IIA fusions with IIA C-terminally linked to EI, but no IIA-EI fusions with IIA N-terminally linked to EI, were identified. Conversely, among tridomain proteins containing EI, HPr and IIA, HPr-I-IIA, and IIA-HPr-I fusions were identified, but no fusions had HPr directly linked to the C terminus of EI. These observations cannot be related to specific stereospecific requirements for the HPr-IIA interaction since both HPr-IIA and IIA-HPr fusion types were found. The explanation may be related to the fact that the HPr binding domain in EI is the N-terminal domain (40, 58). Thus, HPr must be in the proximity of this domain, and consequently, N-terminal but not C-terminal fusions of HPr to EI may be stereospecifically allowed. The fact that IIA can be linked to EI C-terminally but not N-terminally may be similarly explained. Thus, steric hindrance and competition between HPr and IIA may prevent the covalent association of IIA with the N-terminal domain of EI. In other cases where certain domain fusions are favored over others, preferred, but not absolutely required, associative properties of the fused domains may provide an explanation.

We have found PTS protein domains fused to a variety of novel non-PTS proteins and protein domains. The fusion of PTS protein domains to (or within) transcriptional regulators (24, 29, 98, 106) and non-PTS transport proteins (60, 61) had been known previously. Our genome analyses revealed many additional fusions of this general type (Table 13). These included $Na^+/H^+$ antiporter homologues with C-terminally fused $IIA^{Fru}$-like domains, triose-P isomerase homologues with C-terminally fused $IIB^{Glc}$-like domains, and PspF-type putative transcriptional regulatory proteins of the NtrC family, which most closely resemble LevR of *Bacillus subtilis* (5, 16), fused C-terminally to $IIA^{Fru}$-like domains. However, many other types of domains were found in association with PTS protein domains. These included domains homologous to the CBS, MAD, Sbc, HEC1, DUF2, and helix-turn-helix domains. Several of these domains are known to be involved in signal transduction, and more generally, in macromolecular (protein-protein and protein-nucleic acid) interactions. Elucidation of the generalized functions of the non-PTS associative domains will be of great value in determining the specific functions of the fused proteins tabulated in Table 13. Such efforts should keep molecular biologists entertained for decades to come.

## ACKNOWLEDGMENTS

This work was supported by NIH grants GM55434 and GM64368 from the National Institute of General Medical Sciences.

We thank Mary Beth Hiller for her assistance in the preparation of the manuscript.

## REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Austin, S., M. Buck, W. Cannon, T. Eydmann, and R. Dixon.** 1994. Purification and in vitro activities of the native nitrogen fixation control proteins NifA and NifL. J. Bacteriol. **176:**3460–3465.
3. **Bächler, C.** 2004. Catalysis and transcription control by the dihydroxyacetone kinase of *Escherichia coli.* Ph.D. thesis. University of Berne, Berne, Switzerland.
4. **Bächler, C., K. Flükiger-Brühwiler, P. Schneider, P. Bähler, and B. Erni.** 2005. From ATP as substrate to ADP as coenzyme. Functional evolution of the nucleotide binding subunit of dihydroxyacetone kinases. J. Biol. Chem. **280:**18321–18325.
5. **Bao, Q., Y. Tian, W. Li, Z. Xu, Z. Xuan, S. Hu, W. Dong, J. Yang, Y. Chen, Y. Xue, Y. Xu, X. Lai, L. Huang, X. Dong, Y. Ma, L. Ling, H. Tan, R. Chen, J. Wang, J. Yu, and H. Yang.** 2002. A complete sequence of the *T. tengcongensis* genome. Genome Res. **12:**689–700.
6. **Bertram, R., M. Schlicht, K. Mahr, H. Nothaft, M. H. Saier, Jr., and F. Titgemeyer.** 2004. In silico and transcriptional analysis of carbohydrate uptake systems of *Streptomyces coelicolor* A3(2). J. Bacteriol. **186:**1362–1373.
7. **Boël, G., I. Mijakovic, A. Mazé, S. Poncet, M.-K. Taha, M. Larribe, E. Darbon, A. Khemiri, A. Galinier, and J. Deutscher.** 2003. Transcription regulators potentially controlled by HPr kinase/phosphorylase in gram-negative bacteria. J. Mol. Microbiol. Biotechnol. **5:**206–215.
8. **Brown, N. O., J. V. Stoyanov, S. P. Kidd, and J. L. Hobman.** 2003. The MerR family of transcriptional regulators. FEMS Microbiol. Rev. **27:**145–163.

9. **Brüggemann, H., A. Henne, F. Hoster, H. Liesegang, A. Wiezer, A. Strittmatter, S. Hujer, P. Durre, and G. Gottschalk.** 2004. The complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. Science **305:**671–673.
10. **Busch, W. and M. H. Saier, Jr.** 2002. The transporter classification (TC) system, 2002. CRC Crit. Rev. Biochem. Mol. Biol. **37:**287–337.
11. **Canback, B., I. Tamas, and S. G. Andersson.** 2004. A phylogenetic study of endosymbiotic bacteria. Mol. Biol. Evol. **21:**1110–1122.
12. **Cases, I., and V. de Lorenzo.** 2000. Genetic evidence of distinct physiological regulation mechanisms in the $\sigma^{54}$ *Pu* promoter of *Pseudomonas putida.* J. Bacteriol. **182:**956–960.
13. **Chang, A. B., R. Lin, W. K. Studley, C. V. Tran, and M. H. Saier, Jr.** 2004. Phylogeny as a guide to structure and function of membrane transport proteins. Mol. Membr. Biol. **21:**171–181.
14. **Chassy, B. M., and C. A. Alpert.** 1989. Molecular characterization of the plasmid-encoded lactose-PTS of *Lactobacillus casei.*. FEMS Microbiol. Rev. **5:**157–165.
15. **Connelly, J. C., and D. R. Leach.** 1996. The *sbcC* and *sbcD* genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. Genes Cells **1:**285–291.
16. **Debarbouille, M., I. Martin-Verstraete, A. Klier, and G. Rapoport.** 1991. The transcriptional regulator LevR of *Bacillus subtilis* has domains homologous to both sigma 54- and phosphotransferase system-dependent regulators. Proc. Natl. Acad. Sci. USA **88:**2212–2216.
17. **de Crécy-Lagard, V., O. M. M. Bouvet, P. Legeune, and A. Danchin.** 1991. Fructose catabolism in *Xanthomonas campestris* pv. Campestris. Sequence of the PTS operon, characterization of the fructose-specific enzymes. J. Biol. Chem. **266:**18154–18161.
18. **Deutscher, J., B. Pevec, K. Beyreuther, H. H. Kiltz, and W. Hengstenberg.** 1986. Streptococcal phosphoenolpyruvate-sugar phosphotransferase system: amino acid sequence and site of ATP-dependent phosphorylation of HPr. Biochemistry **25:**6543–6551.
19. **Deutscher, J., and M. H. Saier, Jr.** 1983. ATP-dependent protein kinase-catalyzed phosphorylation of a seryl residue in HPr, the phosphoryl carrier protein of the phosphotransferase system in *Streptococcus pyogenes.* Proc. Natl. Acad. Sci. USA **80:**6790–6795.
20. **Doudoroff, M., W. Z. Hassid, E. W. Putman, and A. L. Potter.** 1949. Direct utilization of maltose by *Escherichia coli.*. J. Biol. Chem. **179:**921–934.
21. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J.-F. Tomb, B. A. Dougherty, K. F. Bott, P.-C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchinson III, and J. C. Venter.** 1995. The minimal gene complement of *Mycoplasma genitalium.* Science **270:**397–403.
22. **Garcia-Alles, L. F., C. Siebold, T. L. Nyffeler, K. Flükiger-Bruhwiler, P. Schneider, H. B. Burgi, U. Baumann, and B. Erni.** 2004. Phosphoenolpyruvate- and ATP-dependent dihydroxyacetone kinases: covalent substrate-binding and kinetic mechanism. Biochemistry **43:**13037–13045.
23. **Gonzalez, C. F., A. J. Stonestrom, G. L. Lorca, and M. H. Saier, Jr.** 2005. Biochemical characterization of phosphoryl transfer involving HPr of the phosphoenolpyruvate-dependent phosphotransferase system in *Treponema denticola*, an organism that lacks PTS permeases. Biochemistry **44:**598–608.
24. **Greenberg, D. B., J. Stülke, and M. H. Saier, Jr.** 2002. Domain analysis of transcriptional regulators bearing PTS-regulatory domains. Res. Microbiol. **153:**519–526.
25. **Gutknecht, R., R. Beutler, L. F. Garcia-Alles, U. Baumann, and B. Erni.** 2001. The dihydroxyacetone kinase of *Escherichia coli* utilizes a phosphoprotein instead of ATP as phosphoryl donor. EMBO J. **20:**2480–2486.
26. **Halbedel, S., C. Hames, and J. Stülke.** 2004. In vivo activity of enzymatic and regulatory components of the phosphoenolpyruvate:sugar phosphotransferase system in *Mycoplasma pneumoniae.* J. Bacteriol. **186:**7936–7943.
27. **Higgins, C. F.** 1992. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. **8:**67–113.
28. **Hoischen, C., J. Reizer, A. Dijkstra, S. Rottem, and M. H. Saier, Jr.** 1993. Presence of protein constituents of the gram-positive bacterial phosphotransferase regulatory system in *Acholeplasma laidlawii.* J. Bacteriol. **175:**6599–6604.
29. **Hu, K.-Y., and M. H. Saier, Jr.** 2002. Phylogeny of phosphoryl transfer proteins of the phosphoenolpyruvate-dependent sugar transporting phosphotransferase system. Res. Microbiol. **153:**405–415.
30. **Huang, K., G. Kapadia, P. P. Zhu, A. Peterkofsky, and O. Herzberg.** 1998. A promiscuous binding surface: crystal structure of the IIA domain of the glucose-specific permease from *Mycoplasma capricolum.*. Structure **6:**697–710.
31. **Huber, R., T. A. Langworthy, H. König, M. Thomm, C. R. Woese, U. B. Sleytr, and K. O. Stetter.** 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. Arch. Microbiol. **144:**324–333.
32. **Huber, R., and K. O. Stetter.** 1992. The order *Thermotogales*, p. 3809–3815. *In* A. Balows et al. (ed.), The prokaryotes: a handbook on the biology of bacteria. Ecophysiology, isolation, identification, applications, 2nd ed., vol. 4. Springer, Berlin, Germany.
33. **Hvorup, R. N., A. B. Chang, and M. H. Saier, Jr.** 2003. Bioinformatic analyses of homologues of the bacterial L-ascorbate PTS permeases. J. Mol. Microbiol. Biotechnol. **6:**191–205.
34. **Jessberger, R.** 2002. The many functions of SMC proteins in chromosome dynamics. Nat. Rev. Mol. Cell Biol. **3:**767–778.
35. **King, N. D., and M. R. O'Brian.** 2001. Evidence for direct interaction between enzyme $I^{Ntr}$ and aspartokinase to regulate bacterial oligopeptide transport. J. Biol. Chem. **276:**21311–21316.
36. **Koo, B. M., M. J. Yoon, C. R. Lee, T. W. Nam, Y. J. Choe, H. Jaffe, A. Peterkofsky, and Y. J. Seok.** 2004. A novel fermentation/respiration switch protein regulated by enzyme $IIA^{Glc}$ in *Escherichia coli.* J. Biol. Chem. **279:**31613–31621.
37. **Kristich, C. J., G. D. Glekas, and G. W. Ordal.** 2003. The conserved cytoplasmic module of the transmembrane chemoreceptor McpC mediates carbohydrate chemotaxis in *Bacillus subtilis.* Mol. Microbiol. **47:**1353–1366.
38. **Kundig, W., S. Ghosh, and S. Roseman.** 1964. Phosphate bound to histidine in a protein as an intermediate in a novel phosphotransferase system. Proc. Natl. Acad. Sci. USA **52:**1067–1074.
39. **Lengeler, J. W., and K. Jahreis.** 1996. Phosphotransferase systems or PTSs as carbohydrate transport and as signal transduction systems, p. 573–598. *In* W. N. Konings et al. (ed.), Handbook of biological physics, vol. 2. Elsevier Science B.V., Amsterdam, The Netherlands.
40. **Liao, D. I., E. Silverton, Y. J. Seok, B. R. Lee, A. Peterkofsky, and D. R. Davies.** 1996. The first step in sugar transport: crystal structure of the amino terminal domain of enzyme I of the *E. coli* PEP:sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. Structure **4:**861–872.
41. **Makarova, K. S., L. Aravind, Y. I. Wolf, R. L. Tatusov, K. W. Minton, E. V. Koonin, and M. J. Daly.** 2001. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. Microbiol. Mol. Biol. Rev. **65:**44–79.
42. **Marchler-Bauer, A., and S. H. Bryant.** 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. **32:**W327–W331.
43. **Martin-Lluesma, S., V. M. Stucke, and E. A. Nigg.** 2002. Role of Hec1 in spindle checkpoint signaling and kinetochore recruitment of Mad1/Mad2. Science **297:**2267–2270.
44. **Meier, T. I., R. B. Peery, K. A. McAllister, and G. Zhao.** 2000. Era GTPase of *Escherichia coli*: binding to 16S rRNA and modulation of GTPase activity by RNA and carbohydrates. Microbiology **146:**1071–1083.
45. **Merrick, M. J.** 1993. In a class of its own—the RNA polymerase sigma factor sigma 54 (sigma N). Mol. Microbiol. **10:**903–909.
46. **Merrick, M. J., and J. R. Coppard.** 1989. Mutations in genes downstream of the *rpoN* gene (encoding $\sigma^{54}$) of *Klebsiella pneumoniae* affect expression from $\sigma^{54}$-dependent promoters. Mol. Microbiol. **3:**1765–1775.
47. **Merrick, M. J., and R. A. Edwards.** 1995. Nitrogen control in bacteria. Microbiol. Rev. **59:**604–622.
48. **Merzbacher, M., C. Detsch, W. Hillen, and J. Stülke.** 2004. *Mycoplasma pneumoniae* HPr kinase/phosphorylase. Eur. J. Biochem. **271:**367–374.
49. **Mijakovic, I., S. Poncet, A. Galinier, V. Monedero, S. Fieulaine, J. Janin, S. Nessler, J. A. Marquez, K. Scheffzek, S. Hasenbein, W. Hengstenberg, and J. Deutscher.** 2002. Pyrophosphate-producing protein dephosphorylation by HPr kinase/phosphorylase: a relic of early life? Proc. Natl. Acad. Sci. USA **99:**13442–13447.
50. **Mitchell, W. J., J. Reizer, C. Herring, C. Hoischen, and M. H. Saier, Jr.** 1993. Identification of a phosphoenolpyruvate:fructose phosphotransferase system (fructose-1-P forming) in *Listeria monocytogenes.* J. Bacteriol. **175:**2756–2761.
51. **Moran, N. A.** 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell **108:**583–586.
52. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima.* Nature **399:**323–329.
53. **Nessler, S., S. Fieulaine, S. Poncet, A. Galinier, J. Deutscher, and J. Janin.** 2003. HPr kinase/phosphorylase, the sensor enzyme of catabolite repression in gram-positive bacteria: structural aspects of the enzyme and the complex with its protein substrate. J. Bacteriol. **185:**4003–4010.
54. **O'Toole, R., D. L. Milton, P. Horstedt, and H. Wolf-Watz.** 1997. RpoN of the fish pathogen *Vibrio* (*Listonella*) *anguillarum* is essential for flagellum production and virulence by the water-borne but not intraperitoneal route of inoculation. Microbiology **143:**3849–3859.
55. **Pao, S. S., I. T. Paulsen, and M. H. Saier, Jr.** 1998. The major facilitator superfamily. Microbiol. Mol. Biol. Rev. **62:**1–32.
56. **Paulsen, I. T., J. Reizer, R. Z. Jin, E. C. Lin, and M. H. Saier, Jr.** 2000. Functional genomic studies of dihydroxyacetone utilization in *Escherichia coli.* Microbiology **146:**2343–2344.

57. **Paulsen, I. T., M. K. Sliwinski, and M. H. Saier, Jr.** 1998. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. J. Mol. Biol. **277:**573–592.
58. **Peterkofsky, A., G. Wang, D. S. Garrett, B. R. Lee, Y. J. Seok, and G. M. Clore.** 2001. Three-dimensional structures of protein-protein complexes in the *E. coli* PTS. J. Mol. Microbiol. Biotechnol. **3:**347–354.
59. **Pieper, U., G. Kapadia, P. P. Zhu, A. Peterkofsky, and O. Herzberg.** 1995. Structural evidence for the evolutionary divergence of mycoplasma from gram-positive bacteria: the histidine-containing phosphocarrier protein. Structure **3:**781–790.
60. **Poolman, B., J. Knol, B. Mollet, B. Nieuwenhuis, and G. Sulter.** 1995. Regulation of bacterial sugar-$H^+$ symport by phosphoenolpyruvate-dependent enzyme I/HPr-mediated phosphorylation. Proc. Natl. Acad. Sci. USA **92:**778–782.
61. **Poolman, B., T. J. Royer, S. E. Mainzer, and B. F. Schmidt.** 1989. Lactose transport system of *Streptococcus thermophilus*: a hybrid protein with homology to the melibiose carrier and enzyme III of phosphoenolpyruvate-dependent phosphotransferase systems. J. Bacteriol. **171:**244–253.
62. **Postma, P. W., J. W. Lengeler, and G. R. Jacobson.** 1996. Phosphoenolpyruvate: carbohydrate phosphotransferase systems, p. 1149–1174. *In* F. C. Neidhardt et al. (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D.C.
63. **Powell, B. S., D. L. Court, T. Inada, Y. Nakamura, V. Michotey, X. Cui, A. Reizer, M. H. Saier, Jr., and J. Reizer.** 1995. Novel proteins of the phosphotransferase system encoded within the *rpoN* operon of *Escherichia coli.* J. Biol. Chem. **270:**4822–4839.
64. **Pries, A., H. Priefert, N. Kruger, and A. Steinbuchel.** 1991. Identification and characterization of two *Alcaligenes eutrophus* gene loci relevant to the poly(β-hydroxybutyric acid)-leaky phenotype which exhibit homology to *ptsH* and *ptsI* of *Escherichia coli.* J. Bacteriol. **173:**5843–5853.
65. **Rabus, R., J. Reizer, I. T. Paulsen, and M. H. Saier, Jr.** 1999. Enzyme $I^{Ntr}$ from *Escherichia coli*: a novel enzyme of the phosphoenolpyruvate-dependent phosphotransferase system exhibiting strict specificity for its phosphoryl acceptor, NPr. J. Biol. Chem. **274:**26185–26191.
66. **Reizer, J., C. Hoischen, F. Titgemeyer, C. Rivolta, R. Rabus, J. Stülke, D. Karamata, M. H. Saier, Jr., and W. Hillen.** 1998. A novel protein kinase that controls carbon catabolite repression in bacteria. Mol. Microbiol. **27:** 1157–1169.
67. **Reizer, J., A. Reizer, H. L. Kornberg, and M. H. Saier, Jr.** 1994. Sequence of the *fruB* gene of *Escherichia coli* encoding the diphosphoryl transfer protein (DTP) of the phosphoenolpyruvate:sugar phosphotransferase system. FEMS Microbiol. Lett. **118:**159–162.
68. **Reizer, J., A. Reizer, M. J. Lagrou, K. R. Folger, C. K. Stover, and M. H. Saier, Jr.** 1999. Novel phosphotransferase systems revealed by bacterial genome analysis: the complete repertoire of *pts* genes in *Pseudomonas aeruginosa.* J. Mol. Microbiol. Biotechnol. **1:**289–293.
69. **Reizer, J., A. Reizer, M. J. Merrick, G. Plunkett III, D. J. Rose, and M. H. Saier, Jr.** 1996. Novel phosphotransferase-encoding genes revealed by analysis of the *Escherichia coli* genome: a chimeric gene encoding an enzyme I homologue that possesses a putative sensory transduction domain. Gene **181:**103–108.
70. **Reizer, J., A. Reizer, M. H. Saier, Jr., and G. R. Jacobson.** 1992. A proposed link between nitrogen and carbon metabolism involving protein phosphorylation in bacteria. Protein Sci. **1:**722–726.
71. **Reizer, J., A. Reizer, M. Yamada, and M. H. Saier, Jr.** 1998. The glucitol permease of *Escherichia coli*: a tripartite permease of the phosphotransferase system. Microbiology **144:**1463–1464.
72. **Reizer, J., and M. H. Saier, Jr.** 1997. Modular multidomain phosphoryl transfer proteins of bacteria. Curr. Opin. Struct. Biol. **7:**407–415.
73. **Reizer, J., B. Schneider, A. Reizer, and M. H. Saier, Jr.** 1999. A hybrid response regulator possessing a PEP-dependent phosphorylation domain. Microbiology **145:**987–989.
74. **Reyes-Ramirez, F., R. Little, and R. Dixon.** 2001. Role of *Escherichia coli* nitrogen regulatory genes in the nitrogen response of the *Azotobacter vinelandii* NifL-NifA complex. J. Bacteriol. **183:**3076–3082.
75. **Saier, M. H., Jr.** 1977. Bacterial phosphoenolpyruvate:sugar phosphotransferase systems: structural, functional, and evolutionary interrelationships. Bacteriol. Rev. **41:**856–871.
76. **Saier, M. H., Jr.** 1989. Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Microbiol. Rev. **53:**109–120.
77. **Saier, M. H., Jr.** 2000. A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol. Mol. Biol. Rev. **64:**354–411.
78. **Saier, M. H., Jr., S. Chauvaux, G. M. Cook, J. Deutscher, I. T. Paulsen, J. Reizer, and J.-J. Ye.** 1996. Catabolite repression and inducer control in gram-positive bacteria. Microbiology **142:**217–230.
79. **Saier, M. H., Jr., S. Chauvaux, J. Deutscher, J. Reizer, and J.-J. Ye.** 1995. Protein phosphorylation and the regulation of carbon metabolism in gram-negative versus gram-positive bacteria. Trends Biochem. Sci. **20:**267–271.
80. **Saier, M. H., Jr., B. U. Feucht, and S. Roseman.** 1971. Phosphoenolpyruvate-dependent fructose phosphorylation in photosynthetic bacteria. J. Biol. Chem. **246:**7819–7821.
81. **Saier, M. H., Jr., R. N. Hvorup, and R. D. Barabote.** 2005. Evolution of the bacterial phosphotransferase system: from carriers and enzymes to group translocators. Biochem. Soc. Trans. **33:**220–224.
82. **Saier, M. H., Jr., and M. J. Newman.** 1976. Direct transfer of the phosphoryl moiety of mannitol 1-phosphate to [$^{14}$C] mannitol catalyzed by the enzyme II complexes of the phosphoenolpyruvate:mannitol phosphotransferase systems in *Spirochaeta aurantia* and *Salmonella typhimurium.* J. Biol. Chem. **251:**3834–3837.
83. **Saier, M. H., Jr., M. J. Newman, and A. W. Rephaeli.** 1977. Properties of a phosphoenol-pyruvate:mannitol phosphotransferase system in *Spirochaeta aurantia.* J. Biol. Chem. **252:**8890–8898.
84. **Saier, M. H., Jr., and J. Reizer.** 1994. The bacterial phosphotransferase system: new frontiers 30 years later. Mol. Microbiol. **13:**755–764.
85. **Saier, M. H., Jr., and J. T. Staley.** 1977. Phosphoenolpyruvate:sugar phosphotransferase system in *Ancalomicrobium adetum.* J. Bacteriol. **131:**716–718.
86. **Saier, M. H., Jr., R. Tam, A. Reizer, and J. Reizer.** 1994. Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. Mol. Microbiol. **11:**841–847.
87. **Saier, M. H., Jr., P. Werner, and M. Müller.** 1989. Insertion of proteins into bacterial membranes: mechanism, characteristics and comparisons with the eukaryotic process. Microbiol. Rev. **53:**333–366.
88. **Schmid, K., M. Schupfner, and R. Schmitt.** 1982. Plasmid-mediated uptake and metabolism of sucrose by *Escherichia coli* K-12. J. Bacteriol. **151:**68–76.
89. **Schmidt, A. J., D. A. Ryjenkov, and M. Gomelsky.** 2005. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. J. Bacteriol. **187:**4774–4781.
90. **Segura, D., and G. Espín.** 1998. Mutational inactivation of a gene homologous to *Escherichia coli ptsP* affects poly-β-hydroxybutyrate accumulation and nitrogen fixation in *Azotobacter vinelandii.* J. Bacteriol. **180:**4790–4798.
91. **Seok, Y. J., M. Sondej, P. Badawi, M. S. Lewis, M. C. Briggs, H. Jaffe, and A. Peterkofsky.** 1997. High affinity binding and allosteric regulation of *Escherichia coli* glycogen phosphorylase by the histidine phosphocarrier protein, HPr. J. Biol. Chem. **272:**26511–26521.
92. **Sharma, A., and A. Mondragon.** 1995. DNA topoisomerases. Curr. Opin. Struct. Biol. **5:**39–47.
93. **Siebold, C., I. Arnold, L. F. Garcia-Alles, U. Baumann, and B. Erni.** 2003. Crystal structure of the *Citrobacter freundii* dihydroxyacetone kinase reveals an eight-stranded alpha-helical barrel ATP-binding domain. J. Biol. Chem. **278:**48236–48244.
94. **Siebold, C., L. F. Garcia-Alles, B. Erni, and U. Baumann.** 2003. A mechanism of covalent substrate binding in the X-ray structure of subunit K of the *Escherichia coli* dihydroxyacetone kinase. Proc. Natl. Acad. Sci. USA **100:**8188–8192.
95. **Simm, R., M. Morr, A. Kader, M. Kimtz, and U. Romling.** 2004. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. Mol. Microbiol. **53:**1123–1134.
96. **Simoni, R. D., M. Levinthal, F. D. Kundig, W. Kundig, B. Anderson, P. E. Hartman, and S. Roseman.** 1967. Genetic evidence for the role of a bacterial phosphotransferase system in sugar transport. **58:**1963–1970.
97. **Stonestrom, A., R. D. Barabote, C. Gonzalez, and M. H. Saier, Jr.** 2005. Bioinformatic analyses of bacterial HPr kinase homologues. Res. Microbiol. **156:**443–451.
98. **Stülke, J.** 2002. Control of transcription termination in bacteria by RNA-binding proteins that modulate RNA structures. Arch. Microbiol. **177:**433–440.
99. **Stülke, J., M. Arnaud, G. Rapoport, and I. Martin-Verstraete.** 1998. PRD—a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. Mol. Microbiol. **28:** 865–874.
100. **Stülke, J., and W. Hillen.** 1998. Coupling physiology and gene regulation in bacteria: the phosphotransferase sugar uptake system delivers the signals. Naturwissenschaften **85:**583–592.
101. **Tchieu, J. H., V. Norris, J. S. Edwards, and M. H. Saier, Jr.** 2001. The complete phosphotransferase system in *Escherichia coli.*. J. Mol. Microbiol. Biotechnol. **3:**329–346.
102. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
103. **Titgemeyer, F., K. Jahreis, R. Ebner, and J. W. Lengeler.** 1996. Molecular analysis of the *scrA* and *scrB* genes from *Klebsiella pneumoniae* and plasmid pUR400, which encode the sucrose transport protein enzyme II Scr of the phosphotransferase system and a sucrose-6-phosphate invertase. Mol. Gen. Genet. **250:**197–206.
104. **Titgemeyer, F., J. Walkenhorst, X. Cui, J. Reizer, and M. H. Saier, Jr.** 1994. Proteins of the phosphoenolpyruvate:sugar phosphotransferase system in *Streptomyces*: possible involvement in the regulation of antibiotic production. Res. Microbiol. **145:**89–92.
105. **Ueda, K., A. Yamashita, J. Ishikawa, M. Shimada, T. O. Watsuji, K. Morimura, H. Ikeda, M. Hattori, and T. Beppu.** 2004. Genome sequence of

*Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. Nucleic Acids Res. **32:**4937–4944.

106. **van Tilbeurgh, H., and N. Declerck.** 2001. Structural insights into the regulation of bacterial signalling proteins containing PRDs. Curr. Opin. Struct. Biol. **11:**685–693.
107. **Wassmann, K., V. Liberal, and R. Benezra.** 2003. Mad2 phosphorylation regulates its association with Mad1 and the APC/C. EMBO J. **22:**797–806.
108. **Welch, R. A., V. Burland, G. Plunkett III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Donnenberg, and F. R. Blattner.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli.* Proc. Natl. Acad. Sci. USA **99:**17020–17024.
109. **White, O., J. A. Eisen, J. F. Heidelberg, E. K. Hickey, J. D. Peterson, R. J. Dodson, D. H. Haft, M. L. Gwinn, W. C. Nelson, D. L. Richardson, K. S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J. J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K. S. Makarova, L. Aravind, M. J. Daly, K. W. Minton, R. D. Fleischmann, K. A. Ketchum, K. E. Nelson, S. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science **286:**1571–1577.
110. **Wu, L.-F., J. M. Tomich, and M. H. Saier, Jr.** 1990. Structure and evolution of a multidomain, multiphosphoryl transfer protein: nucleotide sequence of the *fruB*(*HI*) gene in *Rhodobacter capsulatus* and comparisons with homologous genes from other organisms. J. Mol. Biol. **213:**687–703.
111. **Yamada, M., and M. H. Saier, Jr.** 1987. Glucitol-specific enzymes of the phospho-transferase system in *Escherichia coli.* Nucleotide sequence of the *gut* operon. J. Biol. Chem. **262:**5455–5463.
112. **Yu, P. L., R. A. Hodge, and X. P. Li.** 1990. *In vitro* expression of Lac-PTS and tagatose 1,6-bisphosphate aldolase genes from *Lactococcus lactis* subsp. *cremoris* plasmid pDI-21. Appl. Microbiol. Biotechnol. **33:**677–679.
113. **Zhai, Y., J. Tchieu, and M. H. Saier, Jr.** 2002. A web-based Tree View (TV) program for the visualization of phylogenetic trees. J. Mol. Microbiol. Biotechnol. **4:**69–70.
114. **Zuniga, M., I. Comas, R. Linaje, V. Monedero, M. J. Yebra, C. D. Esteban, J. Deutscher, G. Perez-Martinez, and F. Gonzalez-Candelas.** 2005. Horizontal gene transfer in the molecular evolution of mannose PTS transporters. Mol. Biol. Evol. **22:**1673–1685.